Statistics and Computers for Animal and Veterinary Sciences

Fundamentals and Applications

Rakesh Goel

Statistics and Computers for Animal and Veterinary Sciences

About the Author

Rakesh Goel, Incharge, Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, U.P. Pandit Deen Dayal Upadhyaya Pashu Chikitsa Vigyan Vishwavidyalaya Evam Go-Anusandhan Sansthan, Mathura, has been serving the College since 1991. He holds M.Sc. in Statistics (Meerut College, Meerut), M.Phil in Statistics (Meerut University, Meerut). He is associated with the research activities of the students, teachers and research scholars of Veterinary and Animal Science, Animal Biotechnology. He has organized several programmes of Computer training. He has conducted national and regional conventions on the applications of statistics and computers in veterinary sciences.

Statistics and Computers for Animal and Veterinary Sciences

Fundamentals and Applications

Rakesh Goel

Incharge, Department of Animal Genetics and Breeding
College of Veterinary Science and Animal Husbandry
U.P. Pandit Deen Dayal Upadhyay, Pashu Chikitsa Vigyan
Vishwavidyalaya Evam Go Anusandhan Sansthan
Mathura

NEW INDIA PUBLISHING AGENCY
New Delhi – 110 034

NEW INDIA PUBLISHING AGENCY
101, Vikas Surya Plaza, CU Block, LSC Market
Pitam Pura, New Delhi 110 034, India
Phone: + 91 (11)27 34 17 17 Fax: + 91(11) 27 34 16 16
Email: info@nipabooks.com
Web: www.nipabooks.com

Feedback at feedbacks@nipabooks.com

ISBN No. 978-93-86546-55-5 E-ISBN: 978-93-90083-89-3

Composed, Designed & Printed in India

Distributed by NIPA GENX Electronic Resources and Solutions Pvt. Ltd. New Delhi

Dedicated

to

The Sacred Memory

of

My Father, Mother and My Son

Prof. S.K. Garg
Former Vice Chancellor
U.P. Pt. Deen Dayal Upadhyay
Pashu Chikitsa Vigyan Vishwavidyalaya
Evam Go Anusandhan Sansthan
Mathura

FOREWORD

The present book entitled "**Statistics and Computers for Animal and Veterinary Sciences: Fundamentals and Aplications",** has been well prepared to meet the requirements of the undergraduate students of Veterinary and Animal Science, Animal Biotechnology and other related fields.The book has been written to clarify the concepts of statistical methods and methodology of drawing meaningful inferences.

I am sure that the book may serve as a text book not only for students in Veterinary science but also for those who want to know "What statistics in all about" or who need to be familiar with at least the language and fundamental concepts of statistics. The book will serve well to build necessary background for those who will take more advanced courses in statistics including the specialized applications.

I am really very happy to see that the book has been designed in accordance with the new VCI syllabus, 2016 (MSVE-2016). The book will be very useful for students of SAU's/ICAR institutes and those preparing for JRF/SRF/various competitive examinations.

I appreciate the efforts made by the authors in bringing out the book in its present form.

Dated: August, 2017

(S.K. Garg)

Preface

The author feels great pleasure in presenting the book entitled "Statistics and Computers for Animal and Veterinary Sciences: Fundamentals and Aplications". In preparing the book in this field where some books already exist inevitably on wishes to explain why this book is necessary and how it differs from the texts that are already available. The explanation is as follows.

A student or reader of statistics has a large number of books on statistical theory; he can choose according to his needs. But the students of Veterinary and Animal Science, Animal Biotechnology, Medicine, Health and Nutrition is in need of a non-mathematical course on the design and analysis of experiments and the interpretation of the results. This book serves as a great help for this purpose.

In this book, the subject matter has been discussed in such a simple way that the students will find no difficulty to understand it. Each chapter of this book contains complete self explanatory theory and a fairly number of solved examples. We have tried to solve example for each topic in an elegant and more interesting way. Every effort has been made to explain the subject matter in such a simple way that the students can easily understand and feel encouraged to solve questions themselves given in unsolved problems.

It is worthy to record the excellent collaborative efforts and valuable help provided by my reverend colleagues and friends in bringing out this volume.

I am very much hopeful that the present book will be warmly received by the students and teachers. I shall indeed be very thankful to our colleagues for their recommending this book to their students. I lovingly appreciate the tender ideas and help given my wife and dear daughter.

I wish to express my thanks to the publisher for bringing out this book in the present nice form.

Comments and suggestions from the students, researchers and teachers, for the improvement of this book will be highly appreciated and acknowledged.

Rakesh Goel

Contents

Chapter 1

Classification and Tabulation

1.1 Classification

The raw or ungrouped data are in unorganized form and need be organized in meaningful and readily comprehensible form in order to facilitate further statistical analysis. **Classification** of data is the first step in this direction. It is the process of arranging thing in groups according to their resemblance and affinities. There are two basic types of classification namely-**qualitative** and **quantitative**. In **qualitative** classification, the basis of classification is some attribute or quality like: sex, literacy, religion, etc. A qualitative classification may be simple or manifold. The classification done with respect to one attribute is termed as '**Simple Classification'**. The classification in which two or more attributes are considered and several classes are formed is called **'Manifold Classification'**. In **quantitative classification**, the collected data are grouped in respect of characteristics, which can be measured and numerically expressed such as height, weight, etc.

On the basis of these two types of characteristics the classification may be categorized as :

i) Classification according to attributes

ii) Classification according to class-interval

When the given data possess the qualitative characteristic than its classification is known as '**Classification according to attribute'**.

When the given data possess the numerical characteristic to its classification is known as **'Classification according to class-interval'**. For this, we first find out the two extreme values from the given data. The difference of these two highest and lowest observations is known as **Range**. If the number of group or class divides this range we get **width of class or class-interval**. The number of classes depends upon the investigator choice.

$$\text{Width of class or Class interval} = \frac{\text{Range}}{\text{Number of classes}}$$

Generally, number of class is taken 10 and in no case it should exceed to 20. On the bases of class-interval we make different classes so that all the given data may be included in these classes. Each class interval has two i.e. lower and upper limits and for making class interval there are two methods-

i) Inclusive method

ii) Exclusive method

In **inclusive method**, upper limit of the class is not taken equal to lower limit of the next lower class i.e. a-b, c-d, e-f, ... In these class both lower and upper limits are included in same class.

In **exclusive method**, the upper limit of class is taken equal to lower limit of the next higher class i.e. a-b, b-c, c-d, ... In this type, the value equal to upper limit of a class are not included in same class but they are included in the next higher class where it has taken as lower limit.

Division of given data considering the magnitude are put against different class in the form of **Tally mark (/)**. The four tally marks are put parallel to each other and fifth one cross the four (𝍸). This is done actually to facility the counting of like: animals, articles, persons, etc. The total number of tally marks against a class is known as frequency of that class and this is known as **'Frequency Distribution'**. The tabular form of a frequency distribution is called a '**Frequency Table'**. These are two types of frequency tables :

i) Simple frequency table

ii) Cumulative frequency table

In **Simple Frequency Table**, the first column of table contains values of class intervals and second column come their simple frequency.

When frequencies are sum up then they are known as '**Cumulative Frequency'**. There are two types of cumulative frequencies :

i) Less than cumulative frequency

ii) More than cumulative frequency

By adding the frequencies of all observations less than the upper class boundary of a given class, we get **'Less than cumulative frequency'**. Similarly, on adding the frequencies of all observations more than the lower class boundary, one gets **'More than cumulative frequency'**.

1.2 Tabulation of Data

The presentation of classified data in a suitable tabular form is known as "**Tabulation of data**". In other words, tabulation may be defined as the arrangement of data in different rows and columns. The tabulation is always done after classification only. The tabulation is classified as follows-

i) **Simple tabulation:** When sub-division of total data is done on the basis of *one factor* then it is called ***simple tabulation***. It can provide answer of one question only. **For Example**, Division of livestock population on the basis of different states

ii) **Complex tabulation:** It is further sub-divided in three types :

   a) **Double tabulation:** When sub-division of total data is done on the basis of *two factors*, is called ***double tabulation***. It can provide answer of two questions. **For Example**, Division of livestock population first on the basis of states then further sub-division on the basis of species.

   b) **Triple tabulation:** When sub-division of total data is done on the basis of *three factors*, is called ***triple tabulation***. It can provide answer of three questions. **For Example**, Division of livestock population first on the basis of states, second on the basis of species and third on the basis of sex.

   c) **Manifold tabulation:** When sub-division of total data is done on the basis of *more than three factors*, is called ***manifold tabulation***. **For Example**, Division of livestock population first on the basis of states, second on the basis of species, third on the basis of sex and fourth on the basis of maturity i.e. young and adult.

1.2.1 Advantages of tabulation

The advantages of a tabular presentation over the textual presentation are:

i) It is concise

ii) There is no repetition of explanatory matter

iii) Comparisons can be made easily

iv) The important features can be highlighted

v) Errors in the data can be detected.

1.2.2 Preparing a table

An ideal statistical table should contain the following items:

i) **Table number:** A number must be allotted to the table for identification, particularly when there are many tables in a study.

ii) **Title:** The title should explain what is contained in the table. It should be clear, brief and set in bold type on top of the table. It should also indicate the time and place to which the data refer.

iii) **Date:** The date of preparation of the table should be given.

iv) **Stubs or Row designations:** Each row of the table should be given a brief heading. Such designations of rows are called "stubs", or, "stub items" and the entire column is called "stub column".

v) **Column headings or Captions:** Column designation is given on top of each column to explain to what the figures in the column refer. It should be clear and precise. This is called a "caption" or "heading". Columns should be numbered if there are four or more columns.

vi) **Body of the table:** The data should be arranged in such a way that any figure can be located easily. Various types of numerical variables should be arranged in an ascending order, i.e., from left to right in rows and from top to bottom in columns. Column and row totals should be given.

vii) **Unit of measurement:** If the unit of measurement is uniform throughout the table, it is stated at the top right-hand corner of the table along with the title. If different rows and columns contain figures in different units, the units may be stated along with "stubs" or "captions". Very large figures may be rounded up but the method of rounding should be explained.

viii) **Source:** At the bottom of the table a note should be added indicating the primary and secondary sources from which data have been collected.

ix) **Footnotes and references:** If any item has not been explained properly, a separate explanatory note should be added at the bottom of the table.

A table should be logical, well-balanced in length and breadth and the comparable columns should be placed side by side. Light/heavy/thick or double rulings may be used to distinguish sub-columns, main columns and totals. For large data more than one table may be used.

1.2.3 Type of tables

Tables can be classified according to their purpose, stage of enquiry, nature of data or number of characteristics used. On the basis of the number of characteristics, tables may be classified as follows:

i) Simple or one-way table

ii) Two way table

iii) Manifold table

1.2.3.1 Simple or one-way Table

A simple or one-way table is the simplest table which contains data of one characteristic only. A simple table is easy to construct and simple to follow. For example, the blank table given below may be used to show the distribution of animals in different livestock farms in U.P.

Number of animals in different livestock farms in U.P.

Farms	No. of animals
Total	

1.2.3.2 Two-way Table:

A table, which contains data on two characteristics, is called a two-way table. In such case, therefore, either stub or caption is divided into two co-ordinate parts. In the given table, as an example the caption may be further divided in respect of 'species'. This subdivision is shown in two-way table, which now contains two characteristics namely, farms and species.

Sex wise distribution of animals in different livestock farms in U.P.

Farms	No. of animals		Total
	Cattle	Buffalo	
Total			

1.2.3.3 Manifold Table

Thus, more and more complex tables can be formed by including other characteristics. For example, we may further classify the caption sub-headings in the above table in respect of "sex", and "maturity" etc. A table, which has more than two characteristics of data, is considered as a manifold table. For instance, table shown below shows three characteristics namely, farms, species and sex.

Species wise and sex wise distribution of animals in different livestock farms in U.P.

Farms	No. of animals						Total
	Cattle			Buffalo			
	M	F	Total	M	F	Total	
Total							

Foot note: M Stands for Male and F stands for Females.

Manifold tables, though complex are good in practice as these enable full information to be incorporated and facilitate analysis of all related facts. In a normal practice, not more than four characteristics should be represented in one table to avoid confusion. Other related tables may be formed to show the remaining characteristics.

1.3 Frequency Distribution or Grouped Data

If the value of a variable, e.g., height, weight, etc. (continuous), number of students in a class, number of calving in the year at dairy farm (discrete) etc., occurs twice or more in a given series of observations, then the number of occurrence of the value is termed as the "**frequency**" of that value. The way of tabulating a pool of data of a variable and their respective frequencies side by side is called a '**frequency distribution**' of those data. **Croxton and Cowden** defined frequency distribution as "a statistical table which shows the sets of all distinct values of the variable arranged in order of magnitude, either individually or in groups, with their corresponding frequencies side by side".

Let us consider the marks obtained by 50 students of a class in Biostatistics.

Table 1. Marks of 50 Students of a Class in Biostatistics

64	20	71	44	67	31	68	79	78	78
47	50	61	88	35	70	56	89	31	66
58	42	40	37	43	81	84	55	38	29
74	75	25	64	49	51	56	32	45	57
67	55	26	54	63	45	90	63	59	77

If the raw-data of Table 1 are arranged in either ascending, or, descending order of magnitude, we get a better way of presentation, usually called an "array" (Table 2).

Table 2. Array of Marks Shown in Table 2.1

20	31	40	45	54	57	63	67	75	81
25	32	42	47	55	58	64	68	77	84
26	35	43	49	55	59	64	70	78	88
29	37	44	50	56	61	66	71	78	89
31	38	45	51	56	63	67	74	79	90

Now let us present the above data in the form of a simple (or, ungrouped) frequency distribution using the tally marks. A tally mark is an upward slanted stroke **(/)** which is put against a value each time it occurs in the raw data. The fifth occurrence of the value is represented by a cross tally mark **(\)** as shown across the first four tally marks. Finally, the tally marks are counted and the total of the tally marks against each value is its frequency. The total number of tally marks against a class is known as frequency of that class and this is known as **'Frequency Distribution'**. The tabular form of a frequency distribution is called a '**Frequency Table'**. These are two types of frequency tables :

i) **Simple frequency table:** In this type, the first column of table contains values of class intervals and second column come their simple frequency.

ii) **Cumulative frequency table:** When frequencies are sum up then they are known as '**Cumulative Frequency'**. There are two types of cumulative frequencies-

   a) **Less than cumulative frequency:** By adding the frequencies of all observations less than the upper class boundary of a given class, we get **'Less than cumulative frequency'**.

   b) **More than cumulative frequency:** By adding the frequencies of all observations more than the lower class boundary, one gets **'More than cumulative frequency'**.

1.3.1 Formation of discrete frequency distribution

To have a concrete illustration of the construction of a frequency distribution.

Example-1

In a survey of 50 families, the number of buffaloes per family was recorded:

2, 5, 3, 8, 9, 1, 5, 7, 6, 5, 1, 9, 6, 3, 2, 4, 5, 1, 6, 9, 7, 2, 4, 5, 6, 1, 9, 8, 4, 7, 2, 6, 3, 2, 4, 6, 8, 7, 4, 2, 1, 5, 8, 1, 3, 7, 3, 8, 7, 3

Construct frequency distribution and find out the number of families having buffaloes less than 5 and more than 6.

Solution: The following steps are taken.

Step-1: A table with five columns is prepared. In the first column all the values of the variable are written in ascending order starting from lowest to the highest.

Step-2: We go through the values of the given data and insert a tally mark against each value of the variable. If the value again occurs in the data tally mark is again inserted. For the sake of convenience four and cross method, that is after four tallies fifth tally is crossly drawn (~~////~~).

Step-3: We count the number of tallies with respect to each value of the variable and place the third column made for frequency.

Step-4: Less than cumulative frequencies are obtained by adding successively, starting from the top to bottom, simple frequencies. Thus, for calculating these frequencies **'true upper limit'** of a class is regarding as the reference point.

Step-5: More than cumulative frequencies are obtained by adding successively, starting from the bottom to top, simple frequencies. Thus, for calculating these frequencies **'true lower limit'** of a class is regarding as the reference point.

Table 3. Frequency distribution of number of buffaloes per family

No. of buffaloes	Tally Mark	Simple frequency	Less than cumulative frequency	More than cumulative frequency
1	~~////~~, /	6	6	50
2	~~////~~, /	6	12	44
3	~~////~~, /	6	18	38
4	~~////~~	5	**23**	32
5	~~////~~, /	6	29	27
6	~~////~~	5	34	21
7	~~////~~, /	6	40	**16**
8	~~////~~	5	45	10
9	~~////~~	5	50	5
		50		

Result: Hence, the number of families having buffaloes less than 5 is **23** and more than 6 is **16.**

1.3.2 Formation of Continuous Frequency distribution

For a continuous frequency distribution it is necessary to understand the following points.

i) Class limits are the lowest and highest values that can be included in the class.

ii) Class interval is the difference between upper and lower limit of a class. If L = largest, S = smallest value, K = Number of classes then, Class interval, i = (L – S)/K

iii) Mid point is the average of upper (U) and lower (L) limits of the class i.e. Mid-point (m) = (U+L)/2

Example-2

The data collected on weekly milk yield (litres) of 60 Hariana cows are given below :

27	15	24	22	10	35	38	42	21.5	47	49	5
13	10	25	41	39	14	20	1.5	12	2	37	32
46	42	8	7.5	6.5	11	49	48	30	15	5.5	40
1	19	28	24	32	28	34	39	4	45	47	21
2	40	8	28	3	16	41	47	34	32	15	47

Construct a frequency distribution and find out the number of cows giving milk less than 20 litres and more than or equal to 30 litres.

Solution:

Step-1: Determine the largest and smallest value in the given data and calculate the range.

Range = Largest value - Smallest value

Range = 49 - 1 = 48

Step-2: Calculate width of class or class interval

$$\text{Class width or interval (C.I.)} = \frac{\text{Range}}{\text{Number of classes}}$$

Suppose number of classes =10

$$\text{Class width or class interval (C.I.)} = \frac{48}{10} = 4.8 \simeq 5$$

Step-3: Since the smallest value is 1, we start with the lower limit 0 instead of 1, is the first class. The remaining classes can be obtained by adding the class width to each class limit of the previous class until we get the highest class (45-50), which includes the highest value i.e. 49.

Step-4: Less than cumulative frequencies are obtained by adding successively, starting from the top to bottom, simple frequencies. Thus, for calculating these frequencies **'true upper limit'** of a class is regarding as the reference point.

Step-5: More than cumulative frequencies are obtained by adding successively, starting from the bottom to top, simple frequencies. Thus, for calculating these frequencies **'true lower limit'** of a class is regarding as the reference point.

Table 4. Frequency distribution of weekly milk yield (litres)

Weekly milk yield (litres)	Tally Mark	Simple frequency	Less than cumulative frequency	More than cumulative frequency
0-5	𝍸, /	6	6	60
5-10	𝍸, /	6	12	54
10-15	𝍸, /	6	18	48
15-20	𝍸	5	**23**	42
20-25	𝍸, /	6	29	37
25-30	𝍸	5	34	31
30-35	𝍸, /	6	40	**26**
35-40	𝍸	5	45	20
40-45	𝍸, /	6	51	15
45-50	𝍸, ////	9	60	9
		60		

Result: Hence, the number of cows giving milk less than 20 litres is **23** and more than or equal to 30 litres is **26**.

❑❑❑

Chapter 2

Measures of Central Tendency

One of the most important aspects of describing distribution is the central value around which the observations are distributed. A statistical measure used for representing the centre of central value of a set of observations is known as "**Measure of Central Tendency"**. This central value is also called as '**Average'**. The commonly used measures of central tendency are as follows;

i) Arithmetic Mean (A.M.)

ii) Geometric Mean (G.M.)

iii) Harmonic Mean (H.M.)

iv) Weighted Mean (W.M.)

v) Median (M_d)

vi) Mode (M_o)

2.1 Arithmetic Mean (A.M.)

It is a number obtained by dividing the given observations by their number.

2.1.1 Arithmetic mean from individual observations

Following two methods are used for calculating arithmetic mean of an individual series:

a) Direct Method

b) Short cut method

2.1.1.1 Direct Method

If x_1, x_2, x_3, …………….. x_n are n observations, then the arithmetic mean denoted by $\bar{x}$, is given by

$$\overline{X} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum X}{n}$$

2.1.1.2 Short cut Method

If items are more and figures are large enough, the computation of mean becomes difficult. Using the short cut method can solve this difficulty. Under this method an assumed mean is taken as the basis of calculation. The assumed mean is usually chosen to be a neat round number in the middle of the range of the given observations, so that deviations can be easily obtained by subtraction. Then, the formula of short cut method is :

$$\overline{X} = a + \frac{\sum d}{n}$$

Where,

$\overline{X}$ = Arithmetic mean

a = Assumed mean

$\sum d$ = Sum of deviation

Example-1

Calculate arithmetic mean by direct and short-cut method for marks (out of 20) in biostatistics obtained by five students in a class:

Marks: 7 18 12 13 10

Solution:

Using direct method:

Step-1: Add all the individual observations to get Σ X = 60

Step-2: Divide the total number of observations n = 5

Step-3: Calculate arithmetic mean (A.M.) = 60/5 = 12

Now, using short cut method:

Step-1: Taking assumed mean (a) = 10

Step-2: Find deviations of X from assumed mean (a), i.e. d = X –a as follows:

-3, 8, 2, 3, 0

Step-3: Add these deviations to get $\sum d = 10$

Step-4: Calculate arithmetic mean (A.M.) = 10 + (10/5) = 10 + 2 = 12

2.1.2 Arithmetic mean from discrete frequency distribution

2.1.2.1 Direct Method

If the data is given in the form of frequency distribution having values of the variable as $x_1, x_2, x_3, \ldots x_n$ with frequencies $f_1, f_2, f_3, \ldots\ldots,$ f_n respectively. Then, arithmetic mean is defined as :

$$\overline{X} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots + f_nx_n}{f_1 + f_2 + f_3 + \ldots f_n} = \frac{\sum fx}{N}$$

Where,

$N = f_1+f_2+f_3+\ldots\ldots\ldots\ldots+f_n$

2.1.2.2 Short cut Method:

If the data is given in the form of frequency distribution having values of the variable as $x_1, x_2, x_3, \ldots x_n$ with frequencies $f_1, f_2, f_3, \ldots\ldots,$ f_n respectively. Then

$$\overline{X} = a + \frac{\sum fd}{N}$$

Where,

$N = f_1+f_2+f_3+\ldots\ldots\ldots\ldots+f_n$

a = Assumed mean; d = X – a (Deviations of X from a)

Example-2

Calculate Arithmetic mean by direct and short-cut method for body weight (kg) of 40 goats as given in the following frequency table:

Body weight (kg)	:	42	44	46	48	50
No. of goats	:	4	7	15	9	5

Solution:

Using direct method:

Step-1: Prepare following table and find total frequency $\sum f = 40$

Body weight (X)	No. of goats (f)	f.X	d = X –a (Suppose a = 46)	f.d
42	4	168	-4	-16
44	7	308	-2	-14
46	15	690	0	0
48	9	432	2	18
50	5	250	4	20
	N = ∑ f = 40	**1848**		**8**

Step-2: Multiply each value of X with corresponding frequencies and add to get ∑ f.X = 1848

Step-3: Divide the total so obtained by the total frequency to get A.M. = ∑ f.X/N = 1848/40 = 46.20

Now, using short cut method:

Step-1: Taking assumed mean (a) = 46 and obtain deviations of X from a. i.e. d = X -a

Step-2: Multiply each deviation with corresponding frequencies and add to get ∑ f.d = 8

Step-3: Calculate arithmetic mean (A.M.) using formula, A.M. = a + ∑ f.d/N = 46 + (8/40) = 46 + 0.2 = 46.20

2.1.3 Arithmetic mean from continuous frequency distribution

2.1.3.1 Direct Method

If $C_1 - C_2$, $C_2 - C_3$, $C_3 - C_4$, ……….. $C_n - C_{n+1}$ are the class intervals of the data with corresponding frequencies f_1, f_2, f_3, ………, f_n respectively. Then, arithmetic mean is defined as :

$$\overline{X} = \frac{f_1m_1 + f_2m_2 + f_3m_3 + \ldots + f_nm_n}{f_1 + f_2 + f_3 + \ldots f_n} = \frac{\sum fm}{N}$$

Where,

$N = f_1+f_2+f_3+\ldots\ldots\ldots\ldots\ldots+f_n$

m = mid points of class intervals

2.1.3.2 Short cut Method

Let 'a' be assumed mean, d = m – a are deviations of m from assumed mean (a), Then

$$\overline{X} = a + \frac{\sum fd}{N}$$

Where,

N = $f_1+f_2+f_3+\ldots\ldots\ldots\ldots+f_n$

a = Assumed mean; d = m – a (Deviations of m from a)

m = mid points of class intervals

Example-3

Calculate Arithmetic mean by direct and short-cut method for milk production (litres) of 40 cows classified in five groups.

Milk production (Litres):	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of cows :	4	7	15	9	5

Solution:

Using direct method:

Step-1: Prepare following table and find total frequency $\sum$ f = 40

Milk production (Litres)	No. of cows (f)	Mid points (m)	f.m	d = m – a (Suppose a = 25)	f.d
0 – 10	4	5	20	-20	-80
10 – 20	7	15	105	-10	-70
20 – 30	15	25	375	0	0
30 – 40	9	35	315	10	90
40 – 50	5	45	225	20	100
	N = $\sum$ f = 40		**1040**		**40**

Step-2: Multiply each value of m with corresponding frequencies and add to get $\sum$ f.m = 1040

Step-3: Divide the total so obtained by the total frequency to get A.M. = $\sum$ f.m/N = 1040/40 = 26.00

Now, using short cut method:

Step-1: Taking assumed mean (a) = 25 and obtain deviations of m from a. i.e. d = m -a

Step-2: Multiply each deviation with corresponding frequencies and add to get $\sum$ f.d = 40

Step-3: Calculate arithmetic mean (A.M.) using formula, A.M. = a + $\sum$ f.d/N = 25 + (40/40) = 25 + 1.00 = 26.00

2.2 Geometric Mean (G.M.)

Geometric mean (G.M.) is defined to be the n^{th} root of the product of the n quantities of a series.

$$G.M. = (x_1 . x_2 . x_3 . \ldots\ldots\ldots\ldots x_n)^{1/n}$$

2.2.1 Geometric mean from individual observations

if x_1, x_2, x_3, ……….. x_n are the values of n items of the given series. Then, it is defined as-

$$G.M. = (x_1 . x_2 . x_3 . \ldots\ldots\ldots\ldots x_n)^{1/n}$$

When the number of observations is large, the task of determining geometric mean becomes difficult. To simplify calculation, we make use of logarithms and calculate geometric mean from the following formula-

$$\text{Log G.M.} = \frac{1}{n}(\log x_1 + \log x_2 + \log x_3 + \ldots + \log x_n)$$

$$\text{G.M.} = \text{Anti}\log\left[\frac{1}{n}(\log x_1 + \log x_2 + \log x_3 + \ldots + \log x_n)\right]$$

$$\text{G.M.} = \text{Anti}\log\left(\frac{\sum \log x}{n}\right)$$

Example-1

Calculate geometric mean for the daily income of 5 families in a locality is given below:

Daily income (Rs.): 200 225 350 500 400

Solution:

Step-1: Taking log of each given value of X and prepare the following table.

Daily income (X)	Log X
200	2.3010
225	2.3522
350	2.5441
500	2.6990
400	2.6021
	$\sum$ log X = 12.4984

Step-2: Obtain the total $\sum$ log X = 12.4984 and divide the total number of observations n= 5

Step-3: Calculate geometric mean (G.M.) = antilog (12.4984/5) = antilog (2.49968) = 316.00

2.2.2 Geometric mean from discrete frequency distribution

If the data is given in the form of frequency distribution having values of the variable as $x_1, x_2, x_3, ...x_n$ with frequencies $f_1, f_2, f_3, ..., f_n$ respectively. Then, geometric mean is defined as-

$$G.M. = \text{Anti}\log\left(\frac{\sum f \log x}{N}\right)$$

Where,

$N = f_1+f_2+f_3+................+f_n$

Example-2

Calculate geometric mean for body weight (kg) of 40 sheep as given in the following frequency table:

Body weight (kg):	50	54	58	62	66
No. of sheep :	5	8	16	7	4

Solution:

Step-1: Prepare following table and find total frequency $\sum$ f = 40

Body weight (X)	No. of sheep(f)	Log X	f. log X
50	5	1.6990	8.4950
54	8	1.7324	13.8592
58	16	1.7634	28.2144
62	7	1.7924	12.5468
66	4	1.8195	7.2780
	N = $\sum$ f = 40		**70.3934**

Step-2: Multiply these logarithms with corresponding frequencies and add to get $\sum$ f.logX = 70.3934.

Step-3: Divide the total so obtained by the total frequency and take antilog of quotient.

Step-4: Calculate geometric mean (G.M.) using formula, G.M. = antilog ($\sum$ f.logX/N) = antilog (70.3934/40) = antilog (1.7598) = 57.51

2.2.3 Geometric mean from continuous frequency distribution

If $C_1 - C_2$, $C_2 - C_3$, $C_3 - C_4$, ………... $C_n - C_{n+1}$ are the class intervals of the data with corresponding frequencies f_1, f_2, f_3, ………, f_n respectively. Then, geometric mean is defined as :

$$G.M. = \text{Antilog}\left(\frac{\sum f \log m}{N}\right)$$

Where,

$N = f_1+f_2+f_3+$……………..$+f_n$, m = mid-point of different classes

Example-3:

Calculate geometric mean for milk production (litres) of 40 cows classified in five groups.

Milk production (Litres):	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of cows :	4	7	15	9	5

Solution:

Step-1: Find the mid-points (m) of the classes and prepare following table:

Milk production (Litres)	No. of cows (f)	Mid points (m)	Log m	f.log m
0 – 10	4	5	0.6990	2.796
10 – 20	7	15	1.1761	8.2327
20 – 30	15	25	1.3979	20.9685
30 – 40	9	35	1.5441	13.8969
40 – 50	5	45	1.6532	8.266
	N = $\sum$ f = 40			**54.1601**

Step-2: Multiply these logarithms with corresponding frequencies and add to get $\sum$ f.log m = 54.1601.

Step-3: Divide the total so obtained by the total frequency and take antilog of quotient.

Step-4: Calculate geometric mean (G.M.) using formula, G.M. = antilog ($\sum$ f.log m/N) = antilog (54.1601/40) = antilog (1.3540) = 22.59.

2.3 Harmonic Mean (H.M.)

Harmonic mean (H.M.) of a series is the reciprocal of the arithmetic average of the reciprocals of the values of its various items.

$$\text{H.M.} = \frac{n}{\Sigma\left(\frac{1}{x}\right)}$$

2.3.1 Harmonic mean from individual observations

If x_1, x_2, x_3, ……….. x_n are the values of n items of the given series. Then, it is defined as :

$$H.M. = \frac{n}{\left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \ldots + \frac{1}{x_n}\right)}$$

$$H.M. = \frac{n}{\Sigma\left(\frac{1}{X}\right)}$$

Example-1:

Calculate harmonic mean for the daily income of 5 families related to dairy industry in a locality is given below:

Daily income (Rs.):	500	650	750	400	800

Solution:

Step-1: Taking log of each given value of X and prepare the following table:

Daily income (X)	1/X
500	0.0020
650	0.0015
750	0.0013
400	0.0025
800	0.0013
	0.0086

Step-2: Obtain the sum of the reciprocal values i.e. $\sum$ 1/X = 0.0086

Step-3: Calculate harmonic mean (H.M.) = 5/0.0086 = 579.9257

2.3.2 Harmonic mean from discrete frequency distribution

If the data is given in the form of frequency distribution having values of the variable as $x_1, x_2, x_3, ...x_n$ with frequencies $f_1, f_2, f_3, ..., f_n$ respectively. Then, geometric mean is defined as-

$$H.M. = \frac{N}{\Sigma\left(\frac{f}{x}\right)}$$

Where,

$N = f_1+f_2+f_3+................+f_n$

Example-2

Calculate harmonic mean for body weight (kg) of 40 dogs as given in the following frequency table:

Body weight (kg) :	20	24	28	32	36
No. of dogs :	5	8	16	7	4

Solution:

Step-1: Prepare following table and find total frequency $\sum$ f = 40

Body weight (X)	No. of dogs(f)	f/X
20	5	0.2500
24	8	0.3333
28	16	0.5714
32	7	0.2188
36	4	0.1111
	N = $\sum$ f = 40	**1.4846**

Step-2: Divide each frequency by the corresponding value of X and add to get $\sum$ f/X = 1.4846

Step-3: Divide the total frequencies (N = 40) by the total so obtained calculate harmonic mean (H.M.) using formula, H.M. = N/($\sum$ f/X) = 40/1.4846 = 26.9433

2.3.3 Geometric mean from continuous frequency distribution:

If $C_1 - C_2$, $C_2 - C_3$, $C_3 - C_4$, ……….. $C_n - C_{n+1}$ are the class intervals of the data with corresponding frequencies f_1, f_2, f_3, ………, f_n respectively. Then, geometric mean is defined as-

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{f}{m}\right)}$$

Where,

$N = f_1+f_2+f_3+\ldots\ldots\ldots\ldots+f_n$, m = mid-point of different classes

Example-3

Calculate harmonic mean for milk production (litres) of 40 cows classified in five groups.

Milk production (Litres):	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of cows :	4	7	15	9	5

Solution:

Step-1: Find the mid-points (m) of the classes and prepare following table:

Milk production (Litres)	**No. of cows (f)**	**Mid points (m)**	**f/m**
0 – 10	4	5	0.8000
10 – 20	7	15	0.4667
20 – 30	15	25	0.6000
30 – 40	9	35	0.2571
40 – 50	5	45	0.1111
	N = $\sum$ f = 40		**2.2349**

Step-2: Divide each frequencies by the corresponding mid values and add to get $\sum f/m$ = 2.2349

Step-3: Divide the total frequencies (N = 40) by the total so obtained calculate harmonic mean (H.M.) using formula, H.M. = N/($\sum$f /m) = 40/ 2.2349 = 17.8979

2.4 Weighted Mean (W.M.)

The arithmetic mean gives equal importance (weight) to all the observations in a series. Thus weighted mean is used in case when the relative importance of all the items is not equal. Symbolically, if x_1, x_2, x_3, ……….. x_n are the values of n items and w_1, w_2, w_3, ……….. w_n for their respective weights of the given series. Then, it is defined as :

$$W.M. = \frac{(x_1w_1 + x_2w_2 + x_3w_3 + \ldots + x_nw_n)}{w_1 + w_2 + w_3 + \ldots + w_n} = \frac{\sum wx}{\sum(w)}$$

Example-1:

Calculate weighted mean for the performance of the students on the basis of pass percentage and the number of students in each semester in Diploma in Veterinary Pharmacy (DVP) the Veterinary University, Mathura as follows:

Semester	Pass percentage	Number of students
1^{st}	78	70
2^{nd}	62	68
3^{rd}	83	56
4^{th}	78	54

Solution:

Step-1: Prepare the following table and find column totals:

Semester	Pass percentage (X)	Number of students (W)	WX
1^{st}	78	70	5460
2^{nd}	62	68	4216
3^{rd}	83	56	4648
4^{th}	78	54	4212
		$\sum$w = **248**	$\sum$wx =**18536**

Step-2: Multiply each variate value by the corresponding weights and add to get totals $\sum$wx = 18536

Step-3: Divide the total so obtained by the respective sum of the weights to get weighted mean (W.M.) using formula, W.M. = $\sum$wx / $\sum$w = 18536/248 = 74.7419.

2.5 Median (M_d)

Median may be defined as the middle most or central value when the items are arranged in ascending or descending order of magnitude.

2.5.1 Median from individual observations

If $x_1, x_2, x_3, \ldots\ldots\ldots\ldots x_n$ are the values of n items of the given series. Then, in case of **odd number** of items in a data, it is defined as-

$$\text{Median}(M_d) = \left(\frac{n+1}{2}\right)^{th} \text{item}$$

$$\text{Median}(M_d) = \frac{\left(\frac{n}{2}\right)^{th} \text{item} + \left(\frac{n}{2}+1\right)^{th} \text{item}}{2}$$

In case of **even number** of items in a data, then

Example-1

Calculate median for the following daily milk yield (litres) of 7 Sahiwal and 6 Hariana cows respectively.

Daily milk yield of Sahiwal cows (litres): 15 10 8 12 7 11 14 and;
Daily milk yield of Hariana cows (litres): 5 8 4 9 7 4

Solution:

Step-1: Arrange the observations in ascending order of magnitude of first set of cows of Sahiwal breed; i.e. 7, 8, 10, 11, 12, 14, 15

Step-2: Since the number of observations is odd (7), the median is the value of $\{(7+1)/2\}^{th}$ item i.e. $(4)^{th}$ item, Thus, Median = 11 litre

Step-3: Arrange the observations in ascending order of magnitude of second set of cows of Hariana breed; i.e. 4, 4, 5, 7, 8, 9

Step-4: Since the number of observations is even (6), the median is the value of $[(6/2)^{th} + \{(6/2)+1\}^{th}]/2$ item i.e. $\{(3)^{th}$ item + $(4)^{th}$ item$\}/2$, Thus, Median = (5 + 7)/2 = 6 litre

2.5.2 Median from discrete frequency distribution

If the data is given in the form of frequency distribution having values of the variable as $x_1, x_2, \ldots..x_n$ with frequencies $f_1, f_2, f_3, \ldots\ldots\ldots, f_n$ respectively. Then, the procedure of determining median consists of following steps

Step-1: Arrange the data in ascending or descending order of magnitude.

Step-2: Obtain the cumulative frequencies.

Step-3: Determine the size of $\{(N+1)/2\}^{th}$ item. N being the total frequency.

Step-4: Median is located at the value of variable in whose cumulative frequency the value of $\{(N + 1)/2\}^{th}$ item falls.

Example-2

Calculate median for body weight (kg) of 40 dogs as given in the following frequency table:

Body weight (kg)	:	20	24	28	32	36
No. of dogs	:	5	8	16	7	4

Solution:

Step-1: Prepare following table, compute less than type cumulative frequency in the table

Body weight (X)	**No. of dogs (f)**	**Cumulative frequency**
20	5	5
24	8	13
28	16	**29**
32	7	36
36	4	40

Step-2: Calculate Median No. = $\{(N + 1)/2\}^{th}$ value = $\{(40 + 1)/2\}^{th}$ value = $(20.5)^{th}$ value

Step-3: Median is located at the value of the item in whose cumulative frequency the value of $(20.5)^{th}$ item falls. Therefore, Median = 28 kg

2.5.3 Median from continuous frequency distribution

If $C_1 - C_2$, $C_2 - C_3$, $C_3 - C_4$, ……….... $C_n - C_{n+1}$ are the class intervals of the data with corresponding frequencies f_1, f_2, f_3, ………, f_n respectively. Then, the procedure of determining median consists of following steps –

Step-1: Arrange the data in ascending or descending order of magnitude.

Step-2: Obtain the cumulative frequencies.

Step-3: Determine the size of $(N/2)^{th}$ item. N being the total frequency.

Step-4: Locate the median class in cumulative frequency column where the size of $(N/2)^{th}$ item falls.

Step-5: Obtain the median value by applying the formula-

$$\text{Median}\,(M_d) = l_1 + \frac{(l_2 - l_1)(m - c)}{f}$$

Where,

l_1 = Lower limit of median class; l_2 = Upper limit of median class

$m = \frac{N}{2}$; N = Total Frequency

C = Cumulative frequency before the median class; f = Frequency of the median class

Example-3

Calculate median for milk production (litres) of 40 cows classified in five groups.

Milk production (Litres):	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of cows :	4	7	15	9	5

Solution:

Step-1: Prepare following table, compute less than type cumulative frequency in the table

Milk production (Litres)	No. of cows (f)	Cumulative frequency
0-10	4	4
10-20	7	11
20-30	15	26
30-40	9	35
40-50	5	40

Step-2: Calculate Median No. = $(N/2)^{th}$ value = $(40/2)^{th}$ value = $(20)^{th}$ value.

Step-3: Thus, (20 – 30) is the median class. For determine the median in this class, we use the formula :

$$\text{Median}(M_d) = l_1 + \frac{(l_2 - l_1)(m - c)}{f}$$

Here, l_1 = 20, l_2 = 30, m = (N/2) = (40/2) = 20, C = 11, f = 15

$$\text{Median}(M_d) = 20 + \frac{(30 - 20)(20 - 11)}{15} = 20 + \frac{10 \times 9}{15} = 20 + 6 = 26$$

2.6 Mode (M_0)

Mode is that value, which occurs most often in a series. In other words, mode is the most frequently occurring value in a data set.

2.6.1 Mode from individual observations

If x_1, x_2, x_3, ……….. x_n are the values of n items of the given series. Then, for finding mode of a given series count number of times the various values repeat themselves. The value which occur maximum number of times is the mode.

Example-1:

Calculate mode for the following daily milk yield (litres) of 10 Sahiwal cows.

Daily milk yield (litres): 15 10 8 12 7 11 14 12 13 9

Solution:

Step-1: Find the number which occurs most frequently which is 12 occurring 2 times. Thus, Mode = 12 litre.

2.6.2 Mode from discrete frequency distribution

If the data is given in the form of frequency distribution having values of the variable as x_1, x_2, …..x_n with frequencies f_1, f_2, f_3, ………, f_n respectively. Then, mode can be determined by inspection. Here the variate having maximum frequency will be taken as mode.

Example-2:

Calculate mode for body weight (kg) of 5 dogs as given in the following frequency table:

Body weight (kg) :	20	24	28	32	36
No. of dogs :	5	8	16	7	4

Solution:

Step-1: From the above data the variate value 28 has occurred the maximum frequency i.e. 16. Thus, the mode value is 28.

2.6.3 Mode from continuous frequency distribution

If $C_1 - C_2$, $C_2 - C_3$, $C_3 - C_4$, ... $C_n - C_{n+1}$ are the class intervals of the data with corresponding frequencies f_1, f_2, f_3, ..., f_n respectively. Now, a class having maximum frequency is called the modal class. After determine the modal class, the precise value of mode is obtained by using the following formula –

Where,

= Lower limit of modal class;

= Upper limit of modal class

= Frequency of the preceding modal class;

= Frequency of the modal class

= Frequency of the succeeding class

Example-3:

Calculate mode for milk production (litres) of 40 cows classified in five groups.

Milk production (Litres) :	0-10	10-20	20-30	30-40	40-50
No. of cows :	4	7	15	9	5

Solution:

Step-1: Since the highest frequency 15 lies in the class (20 – 30). Therefore, (20 – 30) is the modal class.

Step-2: For determine the mode in this class, we use the formula-

$$\text{Mode}(M_0) = l_1 + \frac{(l_2 - l_1)(f_1 - f_0)}{2f_1 - f_0 - f_2}$$

Here, $l_1 = 20$, $l_2 = 30$, $f_0 = 7$, $f_1 = 15$, $f_2 = 9$

$$\text{Mode}(M_0) = 20 + \frac{(30-20)(15-7)}{2 \times 15 - 7 - 9} = 20 + \frac{(10)(8)}{30 - 7 - 9} = 20 + 5.71 = 25.71$$

❒❒❒

Chapter 3

Measures of Dispersion

The measure of scatteredness of observation around their average is necessary to get a better description of data. The extent or degree to which data tend to spread around an average is called **Dispersion** or **Variation**. **Measure of Dispersion** also called the averages of **second order**; help us to measure the scatteredness of observation around an average. The commonly used measures of dispersion are as follows;

i) Range

ii) Mean deviation (M.D.)

iii) Standard deviation (S.D.)

iv) Variance

v) Standard error (S.E.)

vi) Coefficient of variation (C.V.)

3.1 Range

It is the difference between the lowest and highest value in the series. Thus

Range = L - S

Where,

L = Largest observation

S = Smallest observation

3.1.1 Range from individual observations

If x_1, x_2, x_3, ……….. x_n are the values of n items of the given series. Then, the range is the difference between the largest and the smallest observations.

Example-1

Calculate range for the following daily milk yield (litres) of 7 Sahiwal cows.

Daily milk yield of Sahiwal cows (litres): 15 10 8 12 7 11 14

Solution:

Step-1: The lowest and the highest observations are 7 and 15 respectively. Thus, Range = L – S = 15 – 7 = 8

3.1.2 Range from discrete frequency distribution

If the data is given in the form of frequency distribution having values of the variable as x_1, x_2, ...x_n with frequencies f_1, f_2, f_3, ………, f_n respectively. Then, the range is the difference between the largest and the smallest observations.

Example-2

Calculate range for body weight (kg) of 40 dogs as given in the following frequency table:

Body weight (kg):	20	24	28	32	36
No. of dogs :	5	8	16	7	4

Solution:

Step-1: The lowest and the highest observations are 20 and 36 respectively. Thus, Range = L – S = 36 – 20 = 16.

3.1.3 Median from continuous frequency distribution

If $C_1 - C_2$, $C_2 - C_3$, $C_3 - C_4$, ………... $C_n - C_{n+1}$ are the class intervals of the data with corresponding frequencies f_1, f_2, f_3, ………, f_n respectively. Then, the **range** is the difference between the upper limit of the highest class and the lower limit of the smallest class.

Example-3

Calculate median for milk production (litres) of 40 cows classified in five groups.

Milk production (Litres):	0-10	10-20	20-30	30-40	40-50
No. of cows :	4	7	15	9	5

Solution:

Step-1: The range is the difference between the upper limit of the highest class and the lower limit of the smallest class. Thus, Range = 50 – 0 = 50

3.2 Mean Deviation (M.D.)

It is defined as the arithmetic mean of the absolute values of the deviations of the variant values from their average (Mean or Median or Mode). Symbolically, the **mean deviation** about mean, median or mode can be expressed as follows-

$$\text{Mean deviation about mean} = \frac{\Sigma\left|X - \overline{X}\right|}{n}$$

Where,

$\overline{X}$ = Arithmetic mean

$$\text{Mean deviation about median} = \frac{\Sigma\left|X - M_d\right|}{n}$$

Where,

M_d = Median

$$\text{Mean deviation about mode} = \frac{\Sigma\left|X - M_0\right|}{n}$$

Where,

M_0 = Mode

3.2.1 Mean deviation from individual observations

If x_1, x_2, x_3, ……….. x_n are the values of n items of the given series. Then, mean deviation about mean can be expressed as follows-

$$M.D = \frac{\Sigma\left|X - \overline{X}\right|}{n}$$

Where,

= Arithmetic mean

Example-1

Calculate mean deviation (M.D.) for the following daily milk yield (litres) of 5 Hariana cows.

Daily milk yield of Hariana cows (litres): 8 5 4 7 6

Solution:

Step-1: Prepare the following table:

Milk yield (X)	$\lvert X-\bar{X}\rvert$
8	2
5	1
4	2
7	1
6	0
$\sum$**X = 30**	$\sum\lvert X-\bar{X}\rvert=6$

Step-2: Calculate arithmetic mean (A.M.) = $\sum X/n$ = 30/5 = 6

Step-3: Mean deviation (M.D.) = $\sum\lvert X-\bar{X}\rvert/n$ = 6/5 = 1.20

3.2.2 Mean deviation from discrete frequency distribution

If the data is given in the form of frequency distribution having values of the variable as x_1, x_2, ...x_n with frequencies f_1, f_2, f_3, ..., f_n respectively. Then, the mean deviation (M.D.) about mean can be expressed as follows :

$$M.D.=\frac{\sum f\lvert X-\bar{X}\rvert}{N}$$

Where,

N = Total frequency

Example-2

Calculate mean deviation (M.D.) for body weight (kg) of 40 dogs as given in the following frequency table:

Body weight (kg) :	20	24	28	32	36
No. of dogs :	5	8	16	7	4

Solution:

Step-1: Prepare the following table:

Body weight (X)	**No. of dogs (f)**	**f.X**	$\lvert X-\bar{X}\rvert$	$f\lvert X-\bar{X}\rvert$
20	5	100	7.70	38.5
24	8	192	3.70	29.6
28	16	448	0.30	4.8
32	7	224	4.30	30.1
36	4	144	8.30	33.2
	N = 40	$\sum f$ **.X = 1108**		$\sum f\lvert X-\bar{X}\rvert$ **- % = 136.2**

Step-2: Calculate arithmetic mean (A.M.) = $\sum fx$ /N = 1108/40 = 27.70

Step-3: Mean deviation (M.D.) = $\sum f|X-\overline{X}|$ /N = 136.2/40 = 3.405

3.2.3 Mean deviation from continuous frequency distribution

If $C_1 - C_2$, $C_2 - C_3$, $C_3 - C_4$, ... $C_n - C_{n+1}$ are the class intervals of the data with corresponding frequencies f_1, f_2, f_3, ..., f_n respectively. Then, the mean deviation (M.D.) about mean can be expressed as follows-

$$\text{Mean deviation about mean} = \frac{\sum f|m-\overline{X}|}{N}$$

Where,

N = Total frequency

m = Mid-point of the class interval

Example-3

Calculate mean deviation (M.D.) for milk production (litres) of 40 cows classified in five groups.

Milk production (Litres):	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of cows :	4	7	15	9	5

Solution:

Step-1: Prepare the following table:

Milk	**No. of**	**Mid values**	**f.m**	$|m-\overline{x}|$	$f|m-\overline{x}|$
0 – 10	4	5	20	21	84
10 – 20	7	15	105	11	77
20 – 30	15	25	375	1	15
30 – 40	9	35	315	9	81
40 – 50	5	45	225	19	95
	N = 40		$\sum f$ **.m = 1040**		$\sum f|m-\overline{x}|$ **= 352**

Step-2: Calculate arithmetic mean (A.M.) = $\sum fm$ /N = 1040/40 = 26

Step-3: Mean deviation (M.D.) = $\sum f|m-\overline{x}|$ /N = 352/40 = 8.80

3.3 Standard Deviation (S.D.)

It is defined as the square root of the arithmetic mean of the squares of deviations of the observations from the arithmetic mean. It is denoted by the Greek letter σ (called **Sigma**).

3.3.1 Standard deviation from individual observations

Following two methods are used for calculating **standard deviation** of an individual series:

a) Direct Method

b) Short cut method

3.3.1.1 Direct Method

If x_1, x_2, x_3, …………….. x_n are n observations, then the **standard deviation** denoted by σ is given by

$$\text{S.D.}(\sigma) = \sqrt{\frac{\sum(X-\bar{X})^2}{(n-1)}}$$

Where,

$$\bar{X} = \frac{\sum X}{n}$$

3.3.1.2 Short cut Method

If items are more and figures are large enough, the computation of standard deviation becomes difficult. Using the short cut method can solve this difficulty. Under this method an assumed mean is taken as the basis of calculation. The assumed mean is usually chosen to be a neat round number in the middle of the range of the given observations, so that deviations can be easily obtained by subtraction. Then, the formula of short cut method is-

$$\text{S.D.}(\sigma) = \sqrt{\frac{\sum(X-a)^2 - n(a-\bar{X})}{(n-1)}} \quad X-\bar{X}$$

Example-1:

Calculate standard deviation by direct and short-cut method for marks (out of 10) in biostatistics obtained by five students in a class:

Marks: 7 4 8 5 6

Solution:

Using direct method, prepare the following table:

Marks (X)	$X-\overline{X}$	$(X-\overline{X})^2$
7	1	1
4	-2	4
8	2	4
5	-1	1
6	0	0
$\sum$ **X = 30**		$\sum(X-\overline{X})^2$ **= 10**

Step-1: Calculate arithmetic mean (A.M.) = $\sum X/n$ = 30/5 = 6

Step-2: Calculate standard deviation (S.D.) by the formula :

$$S.D.(\sigma)=\sqrt{\frac{\sum(X-\overline{X})^2}{(n-1)}}=\sqrt{\frac{10}{(5-1)}}=1.58$$

Now, using short cut method:

Step-1: Taking assumed mean (a) = 6

Step-2: Find deviations of X from assumed mean (a), i.e. d = X –a as follows in the table:

Marks (X)	**X – a**	$(X - a)^2$
7	1	1
4	-2	4
8	2	4
5	-1	1
6	0	0
$\sum$ **X = 30**		$\sum(X-a)^2$ **= 10**

Step-3: Calculate standard deviation (S.D.) by the formula :

$$S.D.(\sigma)=\sqrt{\frac{\sum(X-a)^2-n(a-\overline{x})^2}{(n-1)}}=\sqrt{\frac{10-5(6-6)^2}{(5-1)}}=\sqrt{\frac{10-0}{(4)}}=\sqrt{\frac{10}{4}}=1.58$$

3.3.2 Standard deviation from discrete frequency distribution

3.3.2.1 Direct Method

If the data is given in the form of frequency distribution having values of the variable as $x_1, x_2, x_3, \ldots x_n$ with frequencies $f_1, f_2, f_3, \ldots\ldots, f_n$ respectively. Then, **standard deviation** is defined as :

$$S.D.(\sigma)=\sqrt{\frac{\sum f(X-\bar{X})^2}{(N-1)}}$$

Where,

$$\bar{X} = \frac{\sum fX}{N}$$

3.3.2.2 Short cut Method

If the data is given in the form of frequency distribution having values of the variable as $x_1, x_2, x_3, \ldots x_n$ with frequencies $f_1, f_2, f_3, \ldots, f_n$ respectively. Then, **standard deviation** is defined as-

$$S.D.(\sigma) = \sqrt{\frac{\sum f(X-a)^2 - N(a-\bar{X})^2}{(N-1)}}$$

Where,

a = Assumed mean, $N = \sum f = f_1 + f_2 + \ldots\ldots\ldots\ldots\ldots + f_n$

$$\bar{X} = \frac{\sum fX}{N}$$

Example-2:

Calculate standard deviation (S.D.) by direct and short-cut method for body weight (kg) of 40 dogs as given in the following frequency table:

Body weight (kg) :	20	24	28	32	36
No. of dogs :	5	8	16	7	4

Solution:

Step-1: Using direct method, prepare the following table:

Body weight (X)	**No. of dogs (f)**	**f.X**	$X-\bar{X}$	$(X-\bar{X})^2$	$f(X-\bar{X})^2$
20	5	100	-7.70	59.29	296.45
24	8	192	-3.70	13.69	109.52
28	16	448	0.30	0.09	1.44
32	7	224	4.30	18.49	129.43
36	4	144	8.30	68.89	275.56
	N = 40	$\sum f.X = 1108$			$\sum f(X-\bar{X})^2 = 812.40$

Step-2: Calculate arithmetic mean (A.M.) = $\sum fX / N = 1108/40 = 27.70$

Step-3: Calculate standard deviation (S.D.) by the formula :

$$S.D.(\sigma) = \sqrt{\frac{\sum f(X-\overline{X})^2}{(N-1)}} = \sqrt{\frac{812.40}{(40-1)}} = \sqrt{\frac{812.40}{39}} = 4.5641$$

Now, using short cut method:

Step-1: Taking assumed mean (a) = 28

Step-2: Find deviations of X from assumed mean (a), i.e. d = X –a as follows in the table:

Body weight (X)	**No. of dogs (f)**	**f.X**	**X – a**	**(X – a)^2**	**f.(X - a)^2**
20	5	100	-8	64	320
24	8	192	-4	16	128
28	16	448	0	0	0
32	7	224	4	16	112
36	4	144	8	64	256
	N = 40	$\sum f.X$ **= 1108**			$\sum f.(X-a)^2$ **= 816**

Step-3: Calculate standard deviation (S.D.) by the formula :

$$S.D.(\sigma) = \sqrt{\frac{\sum f(X-a)^2 - N(a-\overline{X})^2}{(N-1)}} = \sqrt{\frac{816-40(28-27.70)^2}{(40-1)}} = \sqrt{\frac{812.40}{39}} = \sqrt{20.8308} = 4.5641$$

3.3.3 Standard deviation from continuous frequency distribution

3.3.3.1 Direct Method

If $C_1 - C_2$, $C_2 - C_3$, $C_3 - C_4$, ………... $C_n - C_{n+1}$ are the class intervals of the data with corresponding frequencies f_1, f_2, f_3, ………, f_n respectively. Then, **standard deviation** is defined as :

$$S.D.(\sigma) = \sqrt{\frac{\sum f(m-\overline{X})^2}{(N-1)}}$$

Where,

$$\overline{X} = \frac{\sum fX}{N}$$

3.3.3.2 Short cut Method

Under this method an assumed mean is taken as the basis of calculation. The assumed mean is usually chosen to be a neat round number in the middle of the range of the given observations, so that deviations can be easily obtained by subtraction. Then, the formula of short cut method is-

$$S.D.(\sigma)=\sqrt{\frac{\sum f(m-a)^2-N(a-\overline{X})^2}{(N-1)}}$$

Where, the symbols have usual meaning.

Example-3

Calculate standard deviation (S.D.) for milk production (litres) of 40 cows classified in five groups.

Milk production (Litres)	:	0-10	10-20	20-30	30-40	40-50
No. of cows	:	4	7	15	9	5

Solution:

Step-1: Using direct method, prepare the following table:

Milk Production (CI)	**No. of cows (f)**	**Mid value (m)**	**f.m**	**$m - \overline{X}$**	**$(m - \overline{X})^2$**	**$f(m - \overline{X})^2$**
0-10	4	5	20	-21	441	1764
10-20	7	15	105	-11	121	847
20-30	15	25	375	-1	1	15
30-40	9	35	315	9	81	729
40-50	5	45	225	19	361	1805
	N = 40		**$\sum f.X = 1040$**		**$\sum f.(m - \overline{X})^2 = 5160$**	

Step-2: Calculate arithmetic mean (A.M.) = $\sum fm/N$ = 1040/40 = 26

Step-3: Calculate standard deviation (S.D.) by the formula :

$$S.D.(\sigma)=\sqrt{\frac{\sum f(m-\overline{X})^2}{(N-1)}}=\sqrt{\frac{5160}{(40-1)}}=\sqrt{\frac{5160}{39}}=\sqrt{132.3077}=11.5025$$

Now, using short cut method:

Step-1: Taking assumed mean (a) = 25

Step-2: Find deviations of X from assumed mean (a), i.e. d = X –a as follows in the table:

Milk Production (CI)	No. of cows (f)	Mid value (m)	f.m	m – a	(m – a)^2	f.(m – a)^2
0 – 10	4	5	20	-20	400	1600
10 – 20	7	15	105	-10	100	700
20 – 30	15	25	375	0	0	0
30 – 40	9	35	315	10	100	900
40 – 50	5	45	225	20	400	2000
	N = 40		**$\sum f.X = 1040$**			**$\sum f.(m-\overline{X})^2 = 5200$**

Step-3: Calculate standard deviation (S.D.) by the formula :

$$S.D.(\sigma) = \sqrt{\frac{\sum f(m-a)^2 - N(a-\overline{X})^2}{(N-1)}} = \sqrt{\frac{5200-40(25-26)^2}{(40-1)}} = \sqrt{\frac{5160}{39}} = \sqrt{132.3077} = 11.5025$$

Variance

The square of the standard deviation is known as **variance**. It is denoted by the Greek letter s^2. Thus :

$$\text{Variance} = (S.D.)^2 = \sigma^2 = \frac{\sum(x-\overline{x})^2}{(\eta-1)}$$

Where,

$$\overline{X} = \frac{\sum X}{n}$$

In case of discrete distribution

$$\text{Variance} = (S.D.)^2 = \sigma^2 = \frac{\sum(X-\overline{X})^2}{(N-1)}$$

In case of continuous distribution

$$\text{Variance} = (S.D.)^2 = \sigma^2 = \frac{\sum f(m-\overline{X})^2}{(N-1)}$$

Where,

$$\overline{X} = \frac{\sum fm}{N}$$

$$N = f_1 + f_2 + f_3 + ... + f_n$$

3.4 Standard Error (S.E.)

It is the standard deviation of the statistic concerned in simple sampling. The standard error of a sample mean is given by :

$$S.E. = \frac{\sigma}{\sqrt{n}}$$

In case of grouped data, the formula for calculating **standard error** will be :

$$S.E. = \frac{\sigma}{\sqrt{N}}$$

Where,

$$N = f_1+f_2+f_3+................+f_n$$

3.5 Coefficient of Variation (C.V.)

It is the standard deviation, gives an idea about the extent to which observations are scattered around their mean. Thus, two or more distributions having the same mean can be compared directly for their variability with the help of corresponding standard deviations. The following two situations may arise :

i) When two or more distributions having unequal means are to be compared in respect of their variability.

ii) When two or more distributions having observations expressed in different units of measurements are to be compared in respect of their scatteredness or variability.

For making comparisons in the above two situations, we use a relative measure of dispersion, called **coefficient of variation (C.V.)**. The **coefficient of variation (C.V.)** is defined as :

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Where, the symbols have usual meaning.

❐❐❐

Chapter 4

Diagrammatic Representation of Data

4.1 Diagrammatic Representation of Data

Diagrams are meant only to give a pictorial representation of the data with a view to make them readily intelligible. Several types of diagrams are used for the presentation of data in biological and agriculture sciences. The following are the important types of diagrams in common use :

1. One-dimensional diagrams
2. Two-dimensional diagrams
3. Three-dimensional diagrams

4.1.1 One-dimensional diagrams

They are in the shape of vertical or horizontal lines or bars. The lengths of the lines or bars are in proportion to the different figures they represent. There are three important types of bar diagrams-

i) Simple bar diagrams
ii) Multiple bar diagrams
iii) Sub-divided bar diagrams

4.1.1.1 Simple bar Diagrams

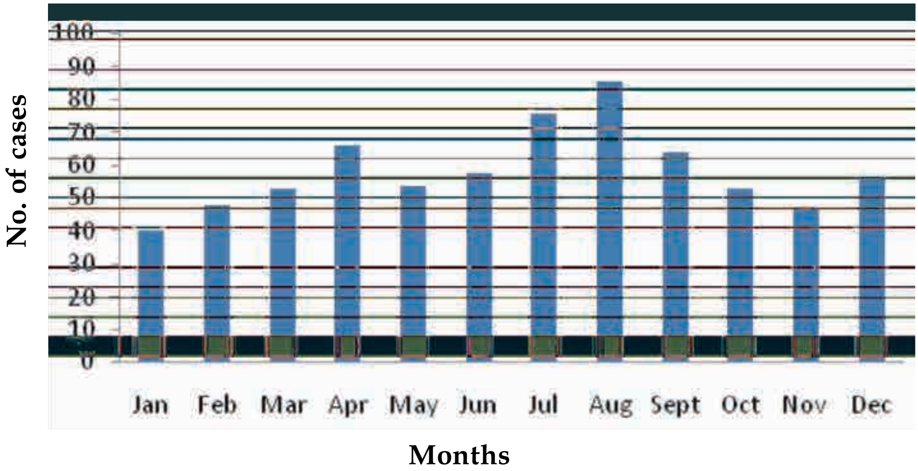
It is constructed by drawing rectangular bars of equal width with different lengths. The bars can be drawn horizontally or vertically. It is used to represent only one variable.

Example-1:

Following are monthly number of cases attended by the doctors in Veterinary Clinical Complex, Veterinary University, Mathura.

Months	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
No. of cases	40	48	53	66	54	58	76	86	64	53	47	56

Simple bar diagram showing number of cases attendant by doctors



4.1.1.2 Multiple bar Diagrams

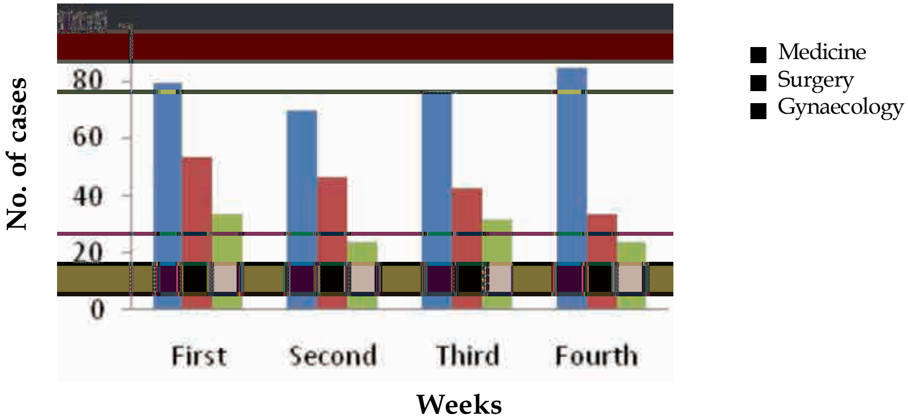
It can represent set of two or more interrelated phenomenon. The method of drawing these diagrams is similar to simple bar diagrams. The only difference is that a separate bar of different shade, colour, etc. represents each characteristic of phenomenon in a particular set. These diagrams are suitable for comparison of the various related phenomenon.

Example-2:

Following are weekly cases attended by the doctors of Medicine, Surgery and Gynaecology in Veterinary Clinical Complex, Veterinary University, Mathura.

Week	**No. of cases attendant by doctors**		
	Medicine	**Surgery**	**Gynaecology**
First	80	54	34
Second	70	47	24
Third	76	43	32
Fourth	85	34	24

Multiple bar diagram showing weekly cases attended by doctors



4.1.1.3 Sub-divided bar Diagrams

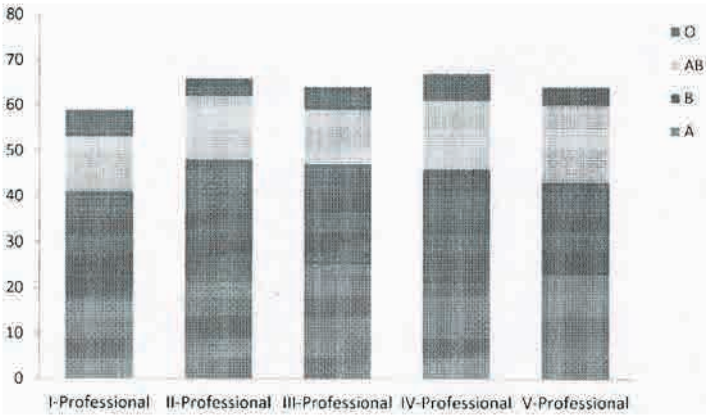
If the given magnitude can be broken into parts of which it is composed or if these are independent quantities constituting sub-division of totals, then the bars may be sub-divided to show the realization of the parts to the whole. In these diagrams, the various components or quantities in each bar should be sub-divided in the same order otherwise proper comparison will not be possible. The components can be distinguished in the bar by different colours or shades.

Example-3:

The following are number of all five professional BVSc & AH students with their blood group in Veterinary University, Mathura.

Blood groups	**No. of BVSc & AH students**				
	I- Profes-sional	**II- Profes-sional**	**III- Profes-sional**	**IV- Profes sional**	**V- Profes-sional**
A	17	21	25	21	23
B	24	27	22	25	20
AB	12	14	12	15	17
O	6	4	5	6	4

Sub-divided bar diagram showing number of students in different blood groups



Blood Group

Professional wise

4.1.2 Two-dimensional diagrams

In these diagrams the length as well as the width of the bars is considered. These are also known as surface diagrams or area diagrams. The important types of such diagrams are :

i) Squares

ii) Circles

iii) Pie diagrams or sector diagrams

iv) Rectangles

4.1.2.1 Squares

In such cases if a bar diagram is drawn then the bars representing big figures would be very big in size while the bars representing the smaller figures would be comparatively very small. In the construction of squares first of all the square root of the various figures is calculated and then squares are drawn with the lengths of their sides in the same proportion as the square root of the original figures. The area of the squares would be in the same proportion as the ratio of original figures.

Example-4:

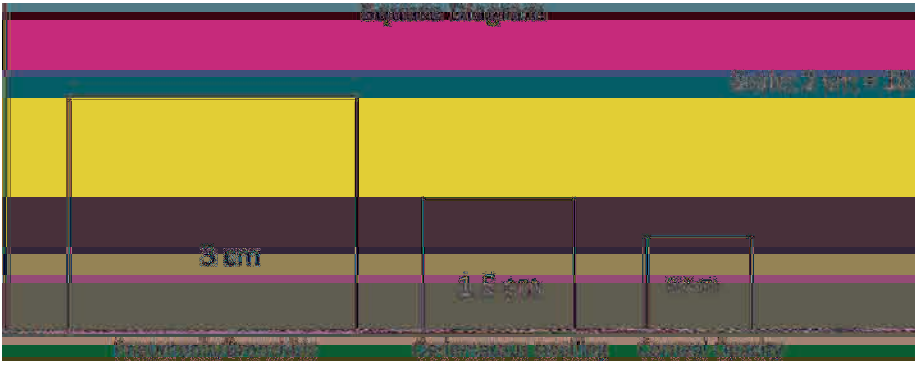
Following data gives the number of affected goat diseases in field flocks in adopted villages in Mathura districts of Uttar Pradesh.

Disease	**Pneumonia/ Bronchitis**	**Oedematous swelling**	**Corneal Opacity**
No. of affected goats	900	225	64

Solution: As there exists a very large proportion between numbers of affected goats, square diagram is suitable. For square diagram, square roots are calculated for the given data value as shown below:

Disease	Pneumonia/ Bronchitis	Oedematous swelling	Corneal Opacity
No. of affected goats	900	225	64
Square root	30	15	8
Sides of square (cm)	3	1.5	0.8

In the above table square root are calculated and each figure has been reduced after division by a similar number say 10. The resultant figures may be taken as the sides of the squares in cm.

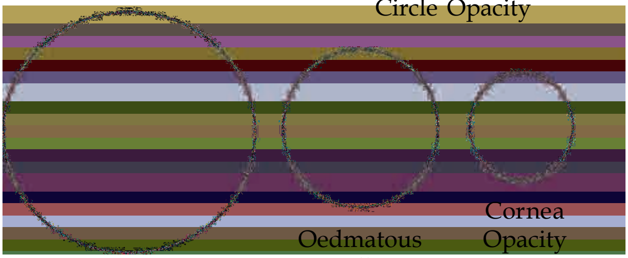


4.1.2.2 Circles

In such diagrams both the total and the component parts can be shown. The area of a circle is proportional to the square of its radius. As in the construction of squares, the square roots of various figures are worked out while constructing the circle. Circles can be used in all those cases in which squares are used. In the above **Example-4**, the square roots are calculated for the given data value as shown below:

Disease	Pneumonia/ Bronchitis	Oedematous swelling	Corneal Opacity
No. of affected goats	900	225	64
Square root	30	15	8
Radius of the circles (cm)	1.5	0.75	0.4

In the above table square root are calculated and each figure has been reduced after division by a similar number say 20. The resultant figures may be taken as the radius of the circles in cm.



4.1.2.3 Pie Diagrams or Sector Diagrams

These diagrams are based on the fact that area of each circle is proportional to its radii. Hence, the radii of the circle represent square root of each magnitude of the data. Also if each magnitude is divided into a number of components, then the 360^0 of the circle are sub-divided according to the proportion of the magnitude of each component, to the total magnitude of the item.

Example-5:

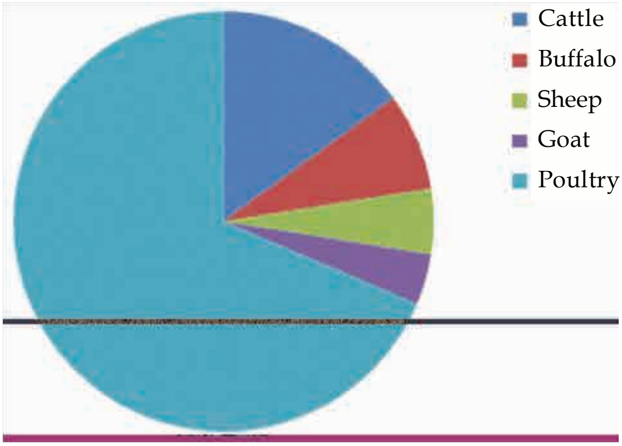
Draw pie diagram for the following data:

Species	**Cattle**	**Buffalo**	**Sheep**	**Goat**	**Poultry**
Livestock population	240	120	80	60	1100

Solution: We will convert these values into corresponding degrees in the circle taking the total percentage 100 as equal to 360^0. The calculation of degrees is shown below:

Species	**Livestock population**	**Angles (in degrees)**
Cattle	240	$\frac{240}{1600} x\ 360^0 = 54$
Buffalo	120	$\frac{120}{1600} x\ 360^0 = 27$
Sheep	80	$\frac{80}{1600} x\ 360^0 = 18$
Goat	60	$\frac{60}{1600} x\ 360^0 = 14$
Poultry	1100	$\frac{1100}{1600} x\ 360^0 = 247$
Total	**1600**	360^0
Square root	**40**	
Radius of circle (cm)	**2 cm**	

Pie Diagram



4.1.2.4 Rectangles

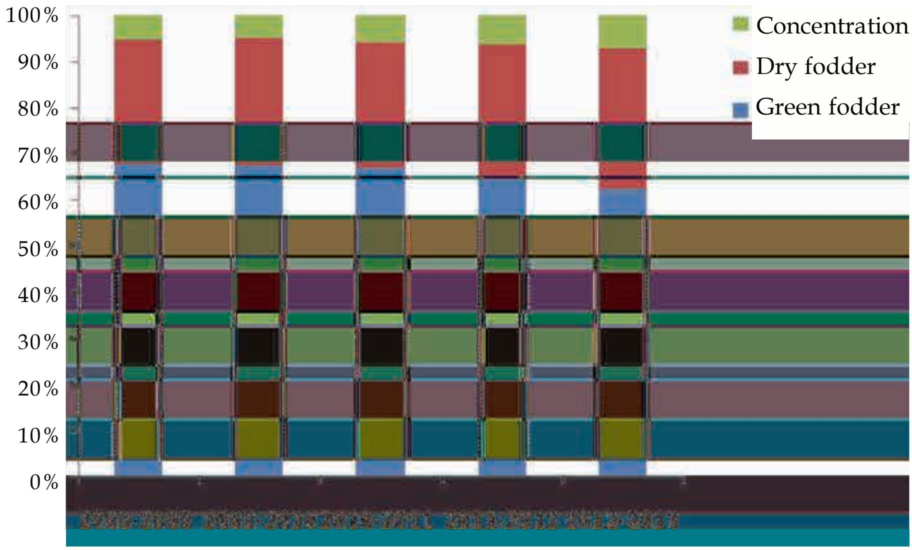
The area of a rectangle is equal to the product of its length and width, while constructing such a diagram both length and width are considered. When two figures are to be shown by the areas of two rectangles, two methods can be adopted, either their width may be kept equal and their lengths in proportion to the two figures or their lengths can be kept equal and their widths in proportion to the size of the two figures. In both the cases the area of the rectangles would be in proportion to the size of the figures.

Example-6:

Following data related to the yearly consumption of ration at Instructional Livestock Farm Complex (ILFC) Veterinary University, Mathura.

Ingredient of Ration	**Yearly Consumption of ration (Quintals)**				
	2008-2009	**2009-2010**	**2010-2011**	**2011-2012**	**2012-2013**
Green fodder	5000	6000	6500	7000	8000
Dry fodder	2000	2500	2700	3200	4000
Concentration	400	450	550	700	900

Rectangular diagram showing consumption of ration



4.1.3 Three-dimensional diagrams

In these diagrams the length, width and height have to be taken into account. Such diagrams are used where the range of difference between the smallest and the largest value is very large i.e. 1:1000. These are also known as volume diagrams. Cylinders, spheres, cubes, etc. are known as three-dimensional diagrams.

Example-7:

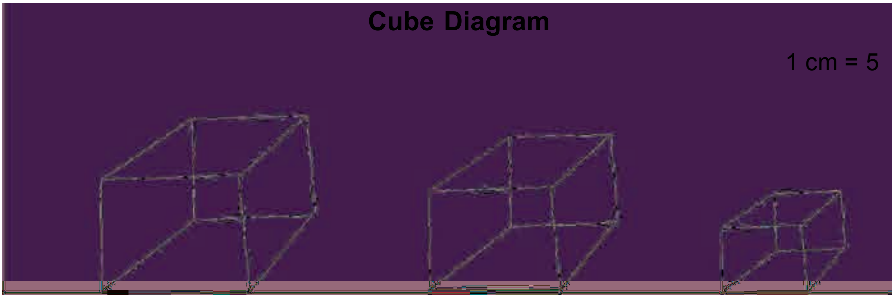
Draw cube diagram for the following data gives the buffalo population in four different countries.

Country	India	Pakistan	Nepal
Buffalo population (thousands)	1000	216	8

Solution: As there exists, a very-very large proportion between buffalo population, cube diagram is suitable. For cube diagram, cube roots are calculated for the given data value as shown below:

Country	India	Pakistan	Nepal
Buffalo population (thousands)	1000	216	64
cube root	10	6	4
Sides of cube (cm)	2	1.2	0.8

In the above table cube root are calculated and each figure has been reduced after division by a similar number say 5. The resultant figures may be taken as the sides of the cubes in cm.



❐❐❐

Chapter 5

Graphical Representation of Data

5.1 Graphical Representation of Data

Diagrams are generally useful for the purpose of publicity and propaganda. No statistical conclusion can be drowning as these provide only an approximate and rough expression about the phenomenon. Thus for a more rigorous and improved representation of numerical statements graphs and charts are used.

Graphs are generally drawn with the help of two perpendicular lines known as x-axis and y-axis. The following are the various types of graphs, which are in common use :

i) Graphs of time series or Historigrams or line graph

ii) Graphs of frequency distribution

5.1.1 Graphs of time series or Historigrams or line graph

The graphs are generally used to represent a data in which the independent variable is time. Here, the time variable is taken on x-axis and the other variable is taken on y-axis. Historigram of two or more variables having same time variant can be plotted on the same graph paper. This makes the comparison easy provided the scale on y-axis for the two variables is same.

Example-1:

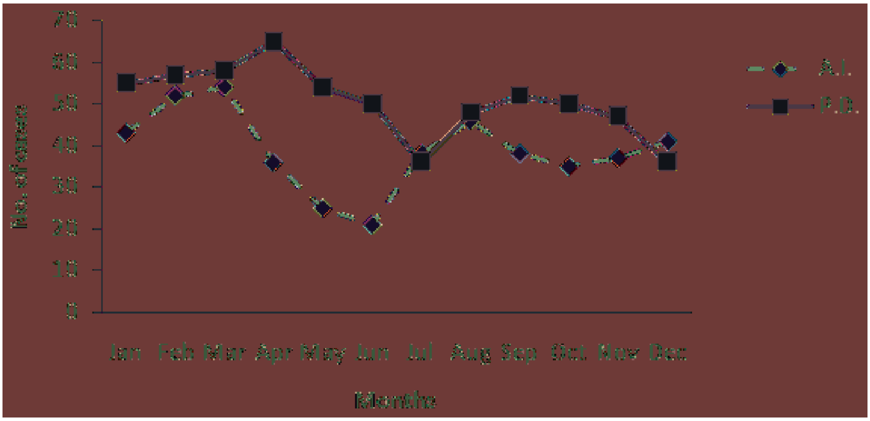
Construct historigram for the following data is related to the number of artificial insemination (A.I.) and pregnancy diagnosis (P.D.) in different months in a year at veterinary clinical complex.

Months:	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
A.I.	43	52	54	36	25	21	38	46	38	35	37	41
P.D.	55	57	58	65	54	50	36	48	52	50	47	36

Solution:

Step-1: For the construction of Historigram, taking time on X-axis and the variables on the Y-axis.

Step-2: Draw two curves by plotted the given data.



5.1.2 Graphs of frequency distribution

Frequency distributions of all types are represented by means of graphs precisely for the same reasons for which graphs are prepared for other types of data. The following types of graphs can be constructed to represent frequency distributions :

i) Histogram

ii) Frequency polygon

iii) Frequency curve

iv) Cumulative frequency curves or Ogives

5.1.2.1 Histogram

It is a graphical representation of the frequency distribution of a continuous variable. A histogram is a set of vertical bars whose areas are proportional to the frequencies represented. In the construction of histogram, class intervals of continuous data are taken on x-axis and the frequencies depending on it on the y-axis. The y-axis represents the frequencies of each class which constitute the height of its rectangles. In this manner, we get a series of rectangles each having a class interval distance as its width and the frequency distance as its height. When class intervals are unequal, a correction for unequal class intervals must be made. The correction consists of finding for each class the frequency density or the relative frequency density. The frequency density is the frequency for that class divided by the width of that class. The total area covered by the histograms for whole frequency distribution is equal to the total number of items in the frequency distribution.

5.1.2.2 Frequency Polygon

Frequency polygon is drawn by joining the consecutive points, plotted by taking mid points of the class intervals on x-axis and corresponding frequencies on y-axis. The end points are extended at each end to join the x-axis. The area of the graph between the x-axis and the two extreme ends of the curve is equal to the total frequency of the data. Frequency polygon can also be drawn with the help of histograms by joining the mid points of the upper side of each rectangle of a histogram.

5.1.2.3 Frequency Curve

The method of drawing a frequency curve is similar to frequency polygon. The only difference is that instead of joining the successive points by straight lines, a free hand smooth curve is drawn in such a way that it passes through most of the points.

Example 2:

Tuberculin reaction measured in 106 persons is as follows:

Reaction (in mm) :	8-10	10-12	12-14	14-16	16-18	18-20	20-22	22-24
No. of person :	5	11	18	26	20	14	8	4

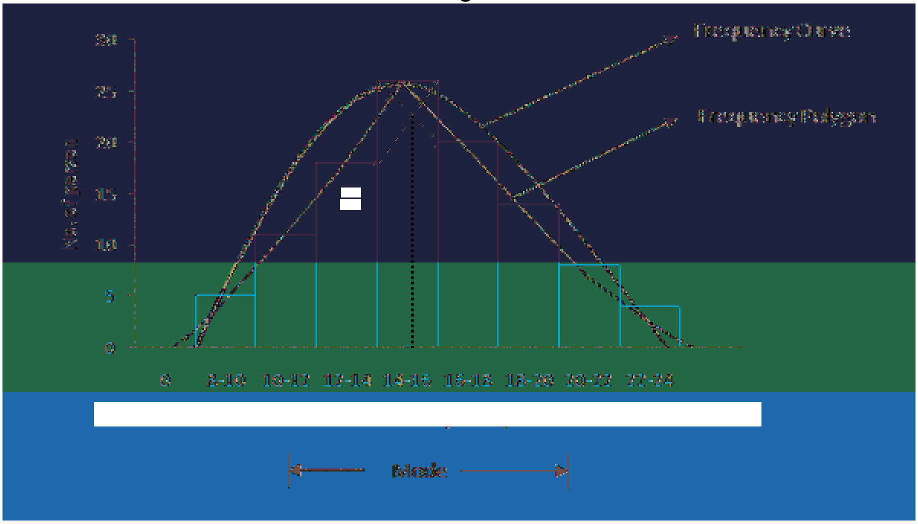
Construct histogram, frequency polygon, frequency curve. Also find the value of mode by graphically.

Solution:

Step-1: For the construction of Histogram, taking tuberculin reaction on X-axis and the variable on the Y-axis.

Step-2: Draw a curve by plotted the given data.

Histogram



5.1.2.4 Cumulative Frequency Curves or Ogives

A graph, where the cumulative frequencies (less than or more than) are plotted against the corresponding class limits (upper or lower) and smoothed out in a free hand curve is known as **cumulative frequency curve or Ogives**. It is known as **more than cumulative frequency curve or Ogive** if more than cumulative frequencies are plotted on the lower limits of the class intervals and **less than cumulative frequency curve or Ogive** if less than cumulative frequencies are plotted on the upper limits of the class intervals. The less than and more **cumulative frequency curve or Ogives** always intersect at **Median**.

Example-3:

Construct cumulative frequency curve for the following frequency distribution:

Milk intake (ml/day):	200-300	300-400	400-500	500-600	600-700	700-800	800-900	900-1000
No. of Students:	7	12	21	26	18	12	9	5

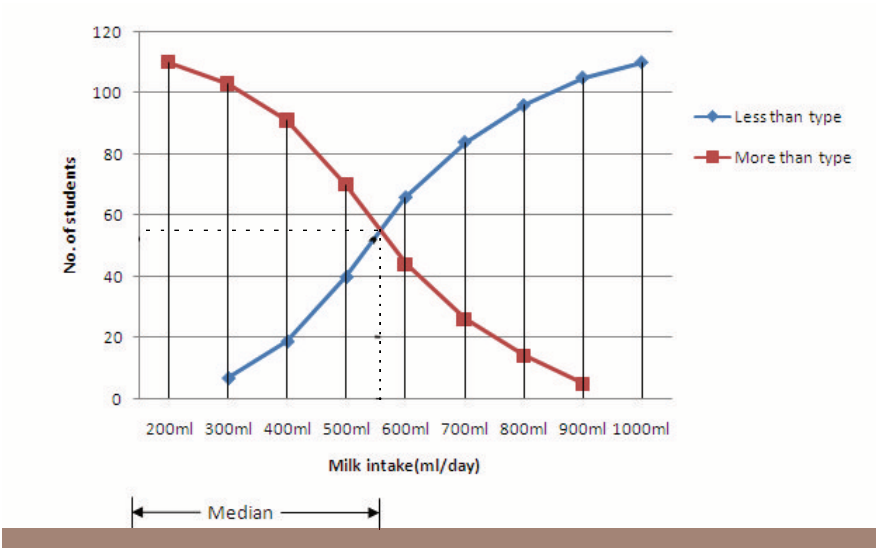
Also find the value of median by graphically.

Solution:

Step-1: For the construction of cumulative frequency curve, first we calculate 'Less than type' and 'More than type' cumulative frequencies and prepare the table:

Milk intake (ml/day)	No. of students (f)	Cumulative frequencies Less than type	More than type
200-300	7	7	110
300-400	12	19	103
400-500	21	40	91
500-600	26	66	70
600-700	18	84	44
700-800	12	96	26
800-900	9	105	14
900-1000	5	110	5

Step-2: Taking milk intake (ml/day) on X-axis and two cumulative frequencies on the Y-axis. Draw a curve by plotted the given data.



❐❐❐

Chapter 6

Correlation and Regression

6.1 Correlation

The term **correlation** indicates the relationship between two such variables in which with changes in the values of one variable, the values of the other variable also change. In short, the tendency of simultaneous variation between two variables is called **correlation**.

6.1.1 Types of correlation

Correlation can be classified by :

i) Positive or negative correlation

ii) Simple, partial and multiple correlations

iii) Linear and non-linear correlation

6.1.1.1 Positive or Negative correlation

The correlation may be classified according to the direction of change in the two variables. In this regard, correlation may be either **positive or negative**.

Positive correlation refers to the movement of variables in the same direction. The correlation is said to be positive when the increase (decrease) in the value of one variable is accompanied by an increase (decrease) in the value of the other variable. In short, two variables move in the same direction are **positive correlation**.

Negative correlation refers to the movement of variables in the opposite direction. The correlation is said to be negative if an increase (decrease) in the value of one variable is followed by decrease (increase) in the value of the other variable. In short, two variables move in the opposite direction are **negative correlation**.

6.1.1.2 Simple, Partial and Multiple correlations

When only two variables are studied it is a problem of **simple correlation**. When three or more variables are studied simultaneously, it is a problem of **multiple correlations**. On the other hand, in **partial correlation** we recognize more than two variables, but consider only two variables to be influencing each other the effect of other influencing variables being kept constant.

6.1.1.3 Linear and Non-linear correlations

In **linear correlation** for every unit change in the values of one variable, there is constant change in the value of other variable. The perfect positive and negative correlations are also linear correlations.

In **non-linear correlation** the change in the values of one variable does not have a constant ratio to the change in the other variable.

6.1.2 Methods of studying correlation

Methods of studying correlation between two variables in the case of ungrouped data may be studies as follows-

i) Scatter diagram

ii) Karl Pearson's coefficient of correlation

iii) Spearman's coefficient of rank correlation

iv) Coefficient of concurrent deviation

v) Method of least squares

6.1.2.1 Scatter Diagram

It is a graphic device for drawing certain conclusions about the correlation between two variables. In preparing a scatter diagram, the observed pairs of observations are plotted on a graph paper in a two dimensional space by taking the measurements on variable x along the horizontal axis and that on the variable y along the vertical axis. The pairs of values are thus represented by dots on the graph. The diagram of dots so obtained is known as **scatter diagram**.

6.1.3 Karl Pearson's coefficient of correlation

Karl Pearson's coefficient of correlation is also known as **coefficient of correlation**. The coefficient of correlation, denoted as r, gives an exact idea about the degree of linear relationship between the two variables x and y and is defined as :

$$r=\frac{\text{Conariance}(x,y)}{\sqrt{\text{Variance}(x),\text{Variance}(y)}}=\frac{\text{Cov}(x,y)}{\sqrt{\text{Var}(x),\text{Var}(y)}}$$

On simplification,

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2,\sum(y-\bar{y})^2}}$$

6.2 Regression

Sir Francis Galton first introduced the word regression, which means reversion, in the study of heredity. The average relationship between two or more variables, which can be used for estimating the value of one variable from the given values of one or more variables, is called **regression**. When only two variables are considered it is termed as **simple regression**. The best average value of one variable associated with the given value of another variable may also be estimated or predicted by means of an equation known as **regression equation**.

6.2.1 Calculation of regression coefficients

The **regression coefficient of y on x**, denoted by b_{yx}, measures the change in the value of y (dependent variable) corresponding to a unit change in the value of x (independent variable). Thus-

$$b_{yx} = \frac{\text{Cov}(x,y)}{\text{Var}(x)} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

Similarly, the regression coefficient of x on y, denoted by b_{xy}, measures the change in the value of x (dependent variable) corresponding to a unit change in the value of y (independent variable). Thus-

$$b_{yx} = \frac{\text{Cov}(x,y)}{\text{Var}(y)} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}$$

6.2.2 Regression lines

If the variables in a bivariate frequency distribution are correlated, we observe that the points in a scatter diagram cluster around a straight called the **regression line**. In a bivariate study, we have two lines of regression, namely :

i) Regression of y on x

ii) Regression of x on y

6.2.2.1 Regression of y on x

The **line of regression of y on x** is used to predict or estimate the value of y for the given value of the variable x. Thus, y is the dependent variable and x is an independent variable. It can be written in the following form :

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

6.2.2.2 Regression of x on y

The **line of regression of x on y** is used to predict or estimate the value of x for the given value of the variable y. In this case, x is the dependent variable and y is an independent variable. It can be written in the following form :

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

6.2.3 Relationship between Regression coefficient and Correlation coefficient:

The correlation coefficient is the geometric mean of the two regression coefficients. Thus :

$$r = \sqrt{b_{yx,b_{xy}}}$$

The sign of the correlation coefficient is the same as that of the two regression coefficients. Thus, r will be positive, if b_{yx} and b_{xy} are positive. Similarly, r will be negative, if b_{yx} and b_{xy} are negative. In this case, this relationship well be written as :

$$r = -\sqrt{b_{yx,b_{xy}}}$$

Example-1:

From the data on age (years) and blood pressure (mm of Hg) of six adult males are given as follows:

Age (years)	:	32	48	38	24	44	34
Blood pressure (mm of Hg)	:	118	128	124	110	124	117

i) Draw scattered diagram and give interpretation of the results.

ii) Calculate correlation coefficient between age and blood pressure.

iii) Calculate regression coefficient of age on blood pressure.

iv) Calculate regression coefficient of blood pressure on age.

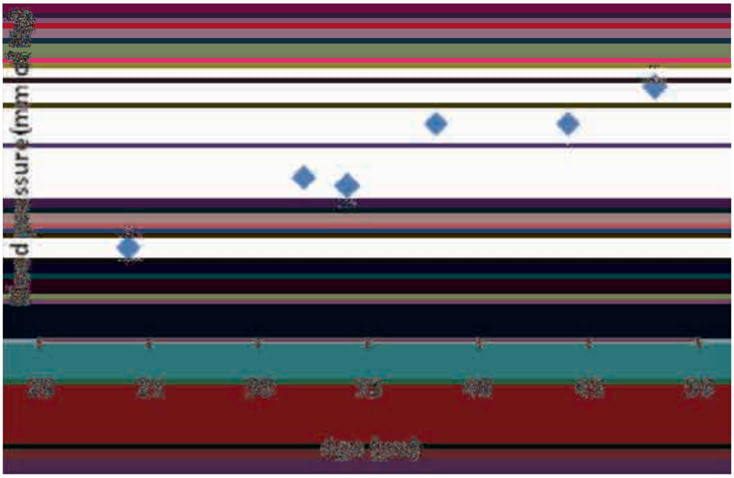
v) Estimate the value of age based on the blood pressure 130 mm.

vi) Estimate the value of blood pressure based at the age of 40 years.

vii) Prove that coefficient of correlation is the geometric mean of two regression coefficients.

Solution:

Step-1: From the scatter diagram as shown below, the plotted points are scattered from left bottom to right top. Therefore, there is positive correlation between age and blood pressure.



Step-2: Prepare the table:

Age(X)	B.P. (Y)	$X - \bar{X}$	$(X - \bar{X})^2$	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
32	118	-5	25	-2	4	10
48	128	11	121	8	64	88
38	124	1	1	4	16	4
26	110	-11	121	-10	100	110
44	124	7	49	4	16	28
34	116	-3	9	-4	16	12
$\sum X = 222$	$\sum Y = 720$		$(X - \bar{X})^2 = 326$		$\sum (Y - \bar{Y})^2 = 216$	$\sum (X - \bar{X})(Y - \bar{Y}) = 252$

Step-3: Calculate arithmetic mean (A.M.) = $\sum X/n$ = 222/6 = 37

Step-4: Calculate arithmetic mean (A.M.) = $\sum Y/n$ = 720/6 = 120

Step-5: Calculate correlation coefficient between age and blood pressure, using the formula:

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2,\sum(y-\bar{y})^2}}=\frac{252}{\sqrt{326-216}}=\frac{252}{\sqrt{70416}}=\frac{252}{265.3601}=0.9497$$

Step-6: Calculate regression coefficient of age on blood pressure, using the formula :

$$b_{xy}=\frac{\mathrm{cov}(x,y)}{Var(y)}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}=\frac{252}{216}=01.1667$$

Step-7: Calculate regression coefficient of blood pressure on age, using the formula :

$$b_{xy}=\frac{\mathrm{cov}(x,y)}{Var(x)}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}=\frac{252}{326}=0.7730$$

Step-8: Estimate the value of age based on the blood pressure 130 mm, using the formula :

$$x-\bar{x}=b_{xy}(y-\bar{y})$$

Here, y = 130mm, then

$$x-37=1.1667(130-120)$$

or,

x = 37 + 11.6670 = 48.6670 years

Step-9: Estimate the value of blood pressure based at the age of 40 years, using the formula :

$$y-\bar{y}=b_{yx}(x-\bar{x})$$

Here, x = 40years, then

$$y-120=0.7730(40-37)$$

or,

y = 120 + 2.3190 = 122.3190 mm

Step-10: Prove that coefficient of correlation is the geometric mean of two regression coefficients, using the formula:

$$r=\sqrt{b_{yx}b_{xy}}=\sqrt{1.1667\times0.7730}=\sqrt{0.9019}=09497$$

❑❑❑

Chapter 7

Probability

7.1 Introduction

The theory of probability has its origin in the games of chance related to gambling such as throwing a die, tossing a coin, drawing cards from a pack of cards etc. **Jerame Cardon** (1501-1576), an Italian mathematician was the first man to write a book on the subject entitled "Book on Games and Chance", which was published after his death in 1663. The systematic and scientific foundation of the mathematical theory of probability was laid in mid-seventeenth century by two French mathematicians **B. Pascal** (1623-1662) and **Pierre de Fermat** (1601-1665).

7.2 Definition

If there are n exhaustive, mutually exclusive and equally likely cases and of them m are favourable to the occurrence of an event A, then the probability of happening of the event A, denoted as P(A), is

$$P(A) = \frac{m}{n} = \frac{Number\ of\ favourable\ cases}{Number\ of\ exhaustive\ cases}$$

and the probability that the event A does not happen will be

$$P(\overline{A}) = \frac{n-m}{n} = \frac{Number\ of\ cases\ unfavourable\ to\ the\ event\ A}{Number\ of\ exhaustive\ cases}$$

Clearly,

$$P(\overline{A}) = 1 - \frac{m}{n} = 1 - P(A)$$

Example-1:

A surgeon transplants the kidney in 200 cases and succeeds in 170 cases. Calculate the probability of success after operation.

Solution:

Step-1: Here, total number of cases (n) = 200

Step-2: Total number of favourable (success) cases (m) = 170

Step-3: By the definition of probability, the probability of success after operation P(A) :

$$P(A) = \frac{m}{n} = \frac{Number\ of\ favourable\ cases}{Number\ of\ exhaustive\ cases} = \frac{170}{200} = 0.85$$

7.3 Theorems of Probability

There are two important theorems of probability :

i) Addition theorem

ii) Multiplication theorem

7.3.1 Addition theorem

This theorem states that if two events A and B are **mutually exclusive**, the probability that any one of them would happen is the sum of the probabilities of the happening of A and B. Symbolically

P(A or B) = P(A) + P(B) or $P(A \cup B) = P(A) + P(B)$

If two events A and B are **not mutually exclusive**, then either A or B both events can occur. In this case, the addition rule is modified as:

P (A or B) = P (A) + P (B) - P (A and B)

or $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example-2:

A card is drawn at random from a well-shuffled pack of 52 cards. What is the probability that it is either a king or a queen?

Solution:

Step-1: Let A be the event of getting a king and B of getting a queen.

Step-2: Now, there are 4 kings and 4 queens in a pack of 52 cards.

Step-3: The probability that it is a king = 4/52

Step-4: The probability that it is a queen = 4/52

Step-5: Since, the events are mutually exclusive, using addition theorem:

Probability that the card drawn is either a king or a queen = P (A or B) = P(A) + P(B) = 4/52 + 4/52 = 2/13.

Example-3:

A card is drawn at random from a well-shuffled pack of 52 cards. What is the probability of getting an ace or a spade?

Solution:

Step-1: Let A be the event of getting an ace and B of getting a spade.

Step-2: Then, A = set of all aces, B = set of all spades and A B = set of an ace of spade.

Step-3: Clearly, n(A) = 4, n(B) = 13 and n(AB) = 1. Also, n(S) =52. Therefore,

Step-4: P(A) = n(A)/n(S) = 4/52, P(B) = n(B)/n(S) = 13/52 and P(AB) = n(AB)/n(S) = 1/52

Step-5: Thus, the required probability:

P (an ace or a spade) = P(A or B) = P(A) + P(B) - P(AB) = 4/52 + 4/52 -1/52 = 4/13

7.3.2 Multiplication theorem

a) When the events are **independent**

This theorem states that if two events A and B are **independent** and can happen simultaneously the probability of their joint occurrence of the event is equal to the product of their separate probabilities. Symbolically

P (A and B) = P (AB) = $P(A \cap B)$ = P (A). P (B)

b) When the events are **dependent**

The probability of simultaneous occurrence of two events A and B is equal to the probabilities of A multiplied by the **conditional probability** of B given A has occurred (or it is equal to the probability of B multiplied by the **conditional probability** of A given that B has occurred). Symbolically

P (A and B) = P (AB) = $P(A \cap B)$ = P (A). P (B/A)

= P (B). P (A/B)

Example-4:

A problem in biostatistics is given to four students say A, B, C and D, their chances of solving it, are 1/2, 1/3, 1/4 and 1/5. What is the probability that the problem will be solved?

Solution:

Step-1: Probability that A fails to solve the problem is 1 – 1/2 = 1/2.

Step-2: Probability that B fails to solve the problem is 1 – 1/3 = 2/3.

Step-3: Probability that C fails to solve the problem is 1 – 1/4 = 3/4.

Step-4: Probability that D fails to solve the problem is 1 – 1/5 = 4/5.

Step-5: Since the events are independent the probability that all the four students fail to solve the problem is:

Step-6: The probability will be solved if anyone of them is able to solve it. Therefore, the probability that the problem will be solved = 1 – 1/5 = 4/5.

Example-5:

A bag contains 4 white and 6 black balls. Two balls are drawn at random one after the other without replacement. Find the probability that both balls drawn are black.

Solution:

Step-1: Probability of drawing a black ball in the first attempt is

$$P(A) = \frac{6}{4+6} = \frac{3}{5}$$

Step-2: Probability of drawing the second black ball given that the first ball drawn is black

$$P(B/A) = \frac{5}{4+5} = \frac{3}{9}$$

Step-3: Therefore, Probability that both balls drawn are black is given by:

$$P(AB) = P(A).P(B/A) = \frac{3}{5} x \frac{5}{9} = \frac{1}{3}$$

❑❑❑

Chapter 8

Normal Distribution

8.1 Introduction

The normal distribution was discovered by **De Moivre** in the year 1733. Normal distribution is probably the most important of all theoretical distributions for the reason that so many physical measurements and natural phenomena have observed frequency distributions which very closely resemble the normal distribution.

8.2 Definition

A continuous random variable X is said to be normally distributed if it has the probability density function represented by the equation-

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

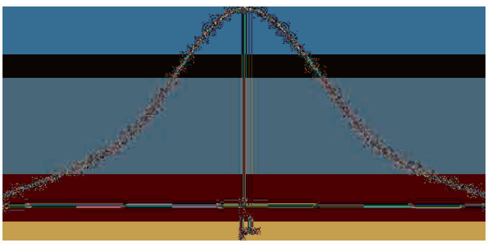
$-\infty < x < \infty$

Where,

μ = Mean of the normal distribution

σ = Standard deviation of the normal distribution.

μ and σ are also known as the two parameters of the normal distribution. The graphical shape of the normal curve is the bell-shaped smooth symmetrical curve.



8.3 Standard Normal Variable (S.N.V.)

It is possible to transform any normal random variable X with mean m and variance s^2 to a new normal random variable Z with mean 0 (zero) and variance 1 (one). This normal random variable Z with mean 0 and variance 1 is called **standard normal variable** (S.N.V.). The transformation of X to Z is :

$$z = \frac{x - \mu}{\sigma}$$

The following important points should be kept in mind while computing area or probability under a standard normal curve :

i) The total area under the standard normal curve is 1.

ii) The mean of the distribution is 0 (zero). Thus, the negative and positive values of Z will lie on the left and right of mean respectively.

iii) The ordinate at mean i.e., at Z = 0 divides the area under the standard normal curve into two equal parts. Thus, the area on the right and left of the ordinate at Z = 0 is 0.5. Symbolically.

$$P(-\infty < Z < 0) = P(0 < Z < \infty) = 0.5$$

iv) Since the curve is symmetrical, thus

$$P(-a < Z < 0) = P(0 < Z \quad a)$$

Example-1:

If X is a normal variable with the mean m = 20 and standard deviation s = 10, find the values of Z_1 and Z_2 such that :

$$P(15 < X < 40) = P(Z_1 < Z < Z_2)$$

Here, Z is a standard normal variable.

Solution:

Step-1: For transforming a normal variable X to a standard normal variable Z, we use the transformation, i.e.

$$z = \frac{x - \mu}{\sigma} = \frac{x - 20}{10}$$

Step-2: Thus, for values of X = 15 and X = 40, the corresponding value of Z variable will be :

When X = 15

$$Z_1 = \frac{X-\mu}{\sigma} = \frac{15-20}{10} = -0.5$$

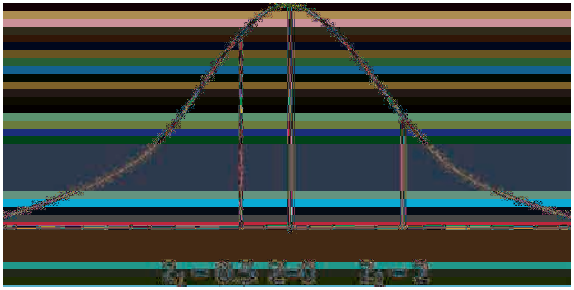
When X = 40

$$Z_2 = \frac{x-\mu}{\sigma} = \frac{40-20}{10} = -2$$

Step-3: Therefore,

P (15 < X < 40) = P (−0.5 < Z < 2)

Step-4: The transformation from X to Z values is also shown in figure



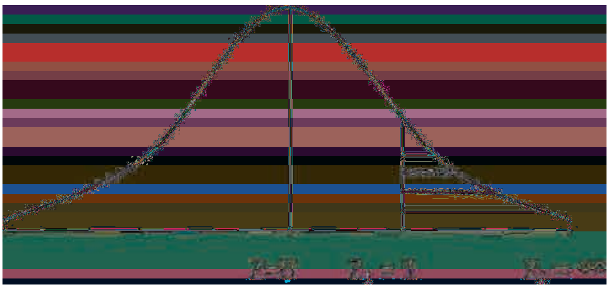
Example-2:

If Z is a standard normal variable. Find the following probabilities or areas

(i) $P(1 < Z < \infty)$ (ii) $P(1 < Z < 2)$

Solution:

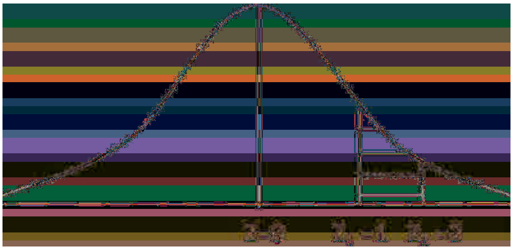
Step-1: (i) A rough sketch of the needed area, as shown by shaded area is helpful in its determination.



Step-2: From this figure, we observe that $P(1 < Z < \infty)$ is the area between the ordinates at $Z_1 = 1$ and $Z_2 = \infty$, thus for $Z_1 = 1$, the required probability is directly determined from table (The Normal Probability Integral table prepared by Fisher & Yates in Statistical table), as :

$P(1 < Z < \infty) = 0.15866$

Step-1: (ii) A rough sketch of the needed area, as shown by shaded area is helpful in its determination.



Step-2: From this figure, we observe that $P(1 < Z < 2)$ is the area between the ordinates at $Z_1 = 1$ and $Z_2 = 2$,

$P(1 < Z < 2) = P(1 < Z < \infty) - P(2 < Z < \infty) = 0.15866 - 0.02275 = 0.13591$

Example-3:

In a distribution exactly normal, 7% of the items are under 35 and 89% are under 63. What are the mean and standard deviation of the distribution?

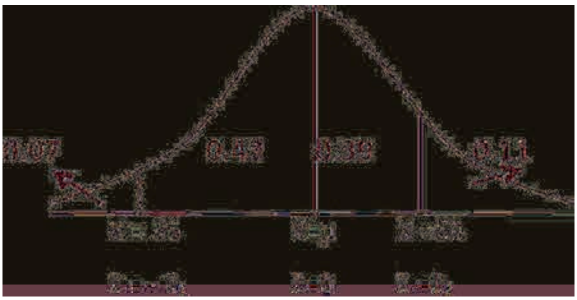
Solution:

Step-1: Let X be the normal variable with mean (μ) and standard deviation (σ). Then we are given that

$P(X < 35) = 0.07$ ...... (i)

$P(X < 63) = 0.89$ ...... (ii)

The locations of the points X = 35 and X = 63 and their corresponding relation with the area is also shown in the figure.



Step-2: Since, the value X = 35 is located to the left of the ordinate at mean, the corresponding Z value will be negative.

Step-3: Thus, for X = 35

$$Z = \frac{35 - \mu}{\sigma} = -Z_{1(say)} \qquad \text{......(iii)}$$

and for X =63

$$Z = \frac{63 - \mu}{\sigma} = -Z_2 \text{ (say)} \qquad \text{...... (iv)}$$

(Z_2 will be positive as it lies to the right of mean)

Step-4: Using the given probabilities in (i) and (ii), we can easily see that

$P(0 < Z < -Z_1) = 0.43$ and $P(0 < Z < Z_2) = 0.39$

Step-5: From table (The Normal Probability Integral table prepared by Fisher & Yates in Statistical table), we get

$Z_1 = 1.48$ and $Z_2 = 1.23$

Step-6: Thus, from (iii) and (iv), one gets

$$\frac{35 - \mu}{\sigma} = -1.48$$

and

$$\frac{63 - \mu}{\sigma} = 1.23$$

Subtracting one gets,

$28/\sigma = 2.71$ or $\sigma = 10.33$

and $\mu = 35 + 1.48 \times 10.33 = 50.30$

Thus, the mean and standard deviation of the normal distribution are 50.30 and 10.33, respectively.

❑❑❑

Chapter 9

Tests of Hypothesis

9.1 Introduction

The tests used to ascertain whether the differences between estimator and the parameter or between two estimators are real or due to chance are called **tests of hypothesis** or **tests of significance**. In other words, the procedures, which enable us to decide whether to accept or reject hypothesis, are called **tests of hypothesis** or **tests of significance**. The following terms are commonly used in testing of hypothesis :

i) Null hypothesis

ii) Alternative hypothesis

9.1.1 Null hypothesis

The hypothesis, which is tested for possible rejection under the assumption that it is true, called **null hypothesis**. It is denoted as H_0.

9.1.2 Alternative hypothesis

Any hypothesis, which is complementary to the null hypothesis, is called an **alternative hypothesis**. It is denoted as H_1.

Example

If we want to test a null hypothesis that the average dry period of Hariana cows in a herd is 150 days, then these two hypothesis can be written as :

H_0: $\mu = 150$ (Null hypothesis)

H_1: $\mu \neq 150$ (Alternative hypothesis)

i.e. $\mu > 150$ or $\mu < 150$

9.2 Level of Significance

In testing of hypothesis, we wish to minimize sizes of both types of errors. However, with fixed size testing procedure, both the errors cannot be minimized simultaneously. Thus, we keep the size or the probability of committing type-I error (α) fixed at certain level, called the **level of significance**. The level of significance is also known as the **size of rejection region** or the **size of the critical region**. The level of significance, which are usually employed in tests of significance are 5% and 1%. If the level of significance is chosen as 5 per cent, it means that the probability of accepting a true hypothesis is 95 per cent.

9.3 Critical Region

A region in the sample space S in which if the computed value of the test statistic lies, we reject the null hypothesis, are called the **critical region** or **rejected region**. When the rejection region consists of two regions each associated with probability $\alpha/2$, called two tailed test. On the other hand, when the rejection region consists of only one region, either on the right or left, associated with probability a, called **one tailed test**.

9.4 Degree of Freedom

It is the number of independent observation used in the making of the statistic. In general, the number of degree of freedom is the total number of observations minus the number of independent constrains imposed on the observations. Thus, if k is the number of independent constraints in a set of data of n observations, then the degree of freedom will be (n - k). The number of degree of freedom for a statistic is usually denoted by λ.

9.5 Procedure of Hypothesis Testing

Following steps can carry out hypothesis testing-

1. Set up the null hypothesis (H_0)
2. Select the level of significance
3. Decide about an appropriate test statistic
4. Find out degree of freedom
5. Find the rejection region and locate the position of the computed test statistic in it. If the test statistic lies in the rejection region, H_0 is rejected otherwise it is not rejected and we conclude that the

data do not provide sufficient evidence to cause rejection of null hypothesis.

6. Finally, we conclude the testing problem with a statistical decision stating clearly the level of significance.

9.6 Large Sample Test (Z-test)

When the samples are of size n >30, almost all distributions are closely approximated by normal distribution, therefore normal distribution form the statistical basis of all the large sample tests. Thus, in the entire large sample test we compute the test statistic Z under H_0 where Z is a standard normal distribution with mean 0 and variance 1. Symbolically, we can Z ~ N(0, 1). The test is usually performed at 5% and 1% level of significance (α = 0.05 and α = 0.01) at which the critical values of Z are 1.96 and 2.58 respectively. We shall discus the following four large sample tests.

i) Testing hypothesis about mean (population variances known)

ii) Testing hypothesis about difference between means (population variances known)

iii) Testing hypothesis of proportion

iv) Testing hypothesis of difference between two proportions

9.6.1 Testing hypothesis about mean (population variances known)

Let x_1, x_2, x_3, …………….. x_n be a random sample of size n drawn from a large population with mean m and known variance s^2. Then, we wish to test

H_0: $\mu = \mu_0$ (Population mean is m_0)

H_1: $\mu \neq \mu_0$ (Population mean differ from m_0)

Where, μ_0 is the specified value of mean.

Test statistic: Under H_0, the test statistic used is Z given by

$$Z = \frac{\overline{X} - \mu_0}{\sigma \sqrt{n}}$$

Here, $\overline{X}$ = sample mean; m_0 = population mean

σ = standard deviation of the population, and

n = sample size

Decision rule: After computing the value of the Z-statistic, the decision about H_0 is taken. If $|Z| \leq 1.96$, we accept H_0 and If $|Z| > 1.96$, we reject H_0, if calculated value of test statistic Z is either greater than 1.96 or less than 1.96 and conclude that the difference between [redacted] and m_0 is significant.

Example-1:

A random sample of 400 flower stems has an average length of 10 cm. Can this be regarded as a sample from a large population with mean of 10.2 cm with a standard deviation 2.25 cm?

Solution:

Step-1: We are given that n = 400, [redacted] μ_0 = 10.2 cm and s = 2.25 cm. Since the sample size n is large, we use Z-statistics. Following steps can carry out hypothesis testing-

Step-2: H_0: There is no significant difference between sample mean [redacted] and population mean μ_0 = 10.2 cm

Step-3: Select the level of significance, generally, used 5% and 1% level of significance.

Step-4: Decide about an appropriate test statistic, i.e.

$$Z = \frac{\overline{X} - \mu_0}{\sigma \sqrt{n}} = \frac{10 - 10.2}{2.25/\sqrt{400}} = \frac{-0.2}{2.25/20} = -1.78$$

Step-5: Here, $|Z_{cal}|$ = 1.78 < 1.96, we accept H_0.

Step-6: Hence, we conclude that there is no significant difference between sample mean and population mean μ_0 = 10.2 cm.

9.6.2 Testing hypothesis about difference between means (population variances known)

Let $x_1, x_2, x_3, \ldots\ldots\ldots\ldots\ldots.. x_{n1}$ be a random sample of size n_1 drawn from a large population with mean m_1 and known variance s_1^2 and $y_1, y_2, y_3, \ldots\ldots\ldots\ldots\ldots.. y_{n2}$ be another random sample of size n_2 drawn from a large population with mean m_2 and known variance s_2^2. Then, we wish to test-

$H_0: \mu_1 = \mu_2$ (No difference between means)

$H_1: \mu_1 \neq \mu_2$ (Significance differ between means)

Test statistic

Under H_0, the test statistic used is Z given by

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Where, and are the means of samples X and Y, respectively.

Remark: If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ i.e. if the samples have been drawn from the same population with common variance σ^2. Then

$$Z = \frac{\overline{X} - \overline{Y}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Decision rule

After computing the value of the Z-statistic, the decision about H_0 is taken. If $|Z| \leq 1.96$, we accept H_0 and If $|Z| > 1.96$, we reject H_0, if calculated value of test statistic Z is either greater than 1.96 or less than 1.96 and conclude that the means of the two population are significant.

Example-2:

A random sample of the height of the 400 boys has a mean of 170 cm and standard deviations of 6.4 cm, while a sample of height of 400 girls has a mean of 160 cm and standard deviations of 6.3 cm. Test whether the girls are on an average taller than the boys?

Solution:

Step-1: H_0: There is no significant difference between the heights of girls and boys.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: Here, $\bar{X} = 170$, $\bar{Y} = 160$, $\sigma_1 = 6.4$, $\sigma_2 = 6.3$, $n_1 = 400$ and $n_2 = 400$, then, decide about an appropriate test statistic, i.e.

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \frac{170 - 160}{\sqrt{\frac{(6.4)^2}{400} + \frac{(6.3)^2}{400}}} = \frac{10}{\sqrt{\frac{40.96}{400} + \frac{39.69}{400}}} = \frac{10}{\sqrt{\frac{80.65}{400}}} = \frac{10}{\sqrt{0.4490}} = 22.27$$

Step-4: Here, $|Z_{cal}|$ = 22.27 > 2.58, then H_0 is rejected at 1% level of significance.

Step-5: Hence, we conclude that there is significant difference between the heights of girls and boys.

9.6.3 Testing hypothesis of proportion

Let p be the proportions of individuals possessing the given attribute in a sample drown from a large population. The null and alternative hypothesis are :

$H_0: p = p_0$ (population proportion is p_0)

$H_1: p \neq p_0$ (population proportion is not p_0)

Where, p_0 is a specific value of population proportion.

Test statistic

Under H_0, the test statistic is :

$$Z = \frac{p - p_0}{\sqrt{\left(\frac{p_0 q_0}{n}\right)}}$$

Where, p_0 = population proportion, $p_0 = 1 - q_0$ and X being the number of individuals possessing in a sample of size n.

Decision rule

After computing the value of the Z-statistic, the decision about H_0 is taken. If $|Z| \leq 1.96$, we accept H_0 and If $|Z| > 1.96$, we reject H_0, if calculated value of test statistic Z is either greater than 1.96 or less than 1.96 and conclude that the sample proportion of attribute in the sample are significant.

Example-3:

Forty people were attacked by a disease and only 36 survived. Can we conclude from sample data that the survival rate, if attacked by the disease, is 85%?

Solution:

Step-1: H_0: the incidence of survival may be taken as 85 percent.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: Here, we are given that

X = No. of survivals = 36; n = No. of persons attacked = 40; therefore, the sample proportion of survivals

p = x/n = 36/40 = 0.90; p_0 = 0.85 and q_0 = 1 - p_0; Now, the test statistic

$$Z=\frac{p-p_0}{\sqrt{\left(\frac{p_0-q_0}{n}\right)}}=\frac{0.90-0.85}{\sqrt{\left(\frac{0.85\times0.15}{40}\right)}}=\frac{0.05}{\sqrt{\left(\frac{0.1275}{40}\right)}}\frac{0.05}{0.056}=0.89$$

Step-4: Here, $|Z_{cal}|$ = 0.89 > 1.96, we accept H_0.

Step-5: Hence, we conclude that the incidence of survival may be taken as 85%.

9.6.4 Testing hypothesis of difference between two proportions

Let p_1 and p_2 are the proportions of individuals possessing the given attribute in random samples of size n_1 and n_2 drown from two large populations. The null and alternative hypothesis are-

H_0: $p_1 = p_0$ (No difference between proportion)

H_1: $p_1 \neq p_0$ (Significant difference in proportion)

Test statistic

Under H_0, the test statistic is :

$$Z=\frac{p_1 - p_2}{\sqrt{\left\{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right\}}}$$

Where, $p_1 = X_1/n_1$, $p_2 = X_2/n_2$, X_1 and X_2 being the number of individuals possessing the given attributes in the samples of size n_1 and n_2, respectively.

p = Combined estimate of proportion or

$$P=\frac{(n_1P_1+n_2p_2)}{(n_1+n_2)}=\frac{(x_1+x_2)}{(n_1+n_2)}$$

q = 1 - p

Decision rule

After computing the value of the Z-statistic, the decision about H_0 is taken. If $|Z| \leq 1.96$, we accept H_0 and If $|Z| > 1.96$, we reject H_0, if calculated value of test statistic Z is either greater than 1.96 or less than 1.96 and conclude that the difference between sample proportions are significant.

Example-4:

In a random sample of 400 persons from city-A, 20 are found to be consumers of meat. In another sample of 300 persons from city-B, 10 are found to be consumers of meat. Do these data reveal a significant difference between city-A and city-B, so far as the proportion of meat consumers is concerned?

Solution:

Step-1: H_0: The proportions of consumers of meat in two cities, say p_1 and p_2 do not differ significantly.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: Here, we are given that, the sample proportion of meat consumers in city-A

$p_1 = x_1/n_1 = 20/400 = 0.05$

Step-4: The sample proportion of meat consumers in city-B

$p_2 = x_2/n_2 = 10/300 = 0.033$

Step-5: The pooled estimate of meat consumers in the two cities is

$p = (n_1p_1 + n_2p_2)/(n_1+n_2) = (x_1+x_2)/ (n_1+n_2) = (20 + 10)/(400 + 300) = 3/70$

therefore, $q = 1 – p = 1 – 3/70 = 67/70$

Now, the test statistic

$$Z = \frac{p_1 - p_2}{\sqrt{\left\{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right\}}} = \frac{0.05 - 0.033}{\sqrt{\left\{\frac{3}{70} x \frac{67}{70}\left(\frac{1}{400}+\frac{1}{300}\right)\right\}}} = \frac{0.017}{0.0154} = 1.103$$

Step-6: Here, $|Z_{cal}| = 1.103 > 1.96$, we accept H_0.

Step-7: Hence, we conclude that the proportions of consumers of meat in two cities.

❐❐❐

Chapter **10**

Small Sample Tests (t-test)

10.1 Introduction

The assumptions on which analysis of large samples is done generally do not hold well in case of small samples ($n < 30$). In case of large samples we had presumed that the random sampling distributions of statistics are approximately normal and further that the values obtained by sampling study are close to the population values and can be used in their place for the calculation of the standard error of the estimate. These assumptions do not hold well if the size of the samples may or may not be normally distributed and similarly it is not possible to substitute the mean or the standard deviation of a small sample in place of the parameter mean or standard deviation for the calculation of standard errors. Under such circumstances the analysis of small samples has to be done by techniques which are different from those applicable in case of large samples.

10.2 Student's t-distribution

This distribution applicable to small samples was developed by **W.S. Gossett** who was employed by the Guinness & Son., Dublin bravery, Ireland. His employer did not permit him to get anything published in his name so he used a pen-name **Student** and get his work published. Therefore, the t-distribution is commonly called **Student's t-distribution**.

If $x_1, x_2, x_3, \ldots\ldots\ldots\ldots\ldots.. x_n$ is a random sample of size n drawn from a normal population with unknown mean m and variance s^2. The t-statistic is defined as :

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{(n-1)d.f.}$$

10.3 Application of t-distribution

The t-distribution has a number of applications of which we will discuss the following:

i) Testing the significance of sample mean (population variance is unknown)

ii) Testing the significance of the difference between two sample means (Unpaired t-test)

iii) Testing the significance of the difference between two means (Paired t-test)

iv) Testing the significance of an observed sample correlation coefficient

v) Testing the significance of an observed sample regression coefficient

10.3.1 Testing the significance of sample mean (population variance is unknown)

Let x_1, x_2, x_3, ... x_n be a random sample of size n drawn from a normal population with a specified mean m and variance s^2. Then, we wish to test

H_0: $\mu = \mu_0$ (Difference between $\bar{x}$ and μ_0 do not differ significantly)

H_1: $\mu \neq \mu_0$ (Difference between and μ_0 differ significantly)

Where, μ_0 is the specified value of mean.

Test statistic

Under H_0, the test statistic used is t, given by

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Here, $\bar{x}$ = sample mean; μ_0 = population mean

n = sample size; s = standard deviation of the sample estimated as

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$$

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (n-1) degree of freedom. After calculating the value of the t-statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner :

i) If the calculated value of test statistic $|t| < t_{(n-1)d.f}(\alpha = 0.05)$, we accept H_0. Hence, we conclude that the difference between $\bar{x}$ and m_0 do not differ significantly.

ii) If the calculated value of test statistic $|t| > t_{(n-1)d.f}(\alpha = 0.05)$, we reject H_0 at 5% level of significance for (n-1) degree of freedom. Hence, we conclude that the difference between $\bar{x}$ and m_0 differ significantly.

iii) If the calculated value of test statistic $|t| > t_{(n-1)d.f}(\alpha = 0.01)$, we reject H_0 at 1% level of significance for (n-1) degree of freedom. Hence, we conclude that the difference between $\bar{x}$ and m_0 **highly** differ significantly.

Example-1:

A random sample of 7 cows selected from a large population gave the following data on life time milk production:

Production (0000 litre): 5.2 3.9 5.6 4.1 5.2 5.8 5.2

Test whether the average life time production in the population is 40000 litres?

Solution:

Step-1: H_0: There is no significance difference between sample mean with the population is 40000 litres.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Step-4: For using test statistic, prepare the following table

X	$X - \bar{X}$	$(X - \bar{X})^2$
5.2	0.2	0.04
3.9	-1.1	1.21
5.6	0.6	0.36
4.1	-0.9	0.81
5.2	0.2	0.04
5.8	0.8	0.64
5.2	0.2	0.04
$\sum X$ **= 35**		$(X - \bar{X})^2$ **= 3.14**

$$\overline{X} = \frac{\sum X}{n} = \frac{35}{7} = 5$$

Step-5: Calculate standard deviation of sample,

$$s = \sqrt{\frac{\sum(X-\overline{X})^2}{n-1}} = \sqrt{\frac{3.14}{7-1}} = \sqrt{0.52} = 0.72$$

Now, the test statistic

$$t = \frac{5-4}{0.72/\sqrt{7}} = \frac{1}{0.72/2.65} = 3.68$$

Step-6: Degree of freedom = n – 1 = 7 – 1 = 6

Step-7: Table value of t (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$t_6(0.05) = 2.45$ and $t_6(0.01) = 3.71$

Step-8: Here, $|t_{cal}| = 3.68 < t_{tab} = 2.45$, we reject H_0 at 5% level of significance.

Step-9: There is significance difference between sample mean with the population is 40000 litres.

10.3.2 Testing the significance of the difference between two sample means (Unpaired t-test)

Let $x_1, x_2, x_3, \ldots\ldots\ldots\ldots\ldots.. x_{n1}$ be a random sample of size n_1 drawn from a normal population with mean m_1 and unknown variance s_1^2 and $y_1, y_2, y_3, \ldots\ldots\ldots\ldots\ldots.. y_{n2}$ be another random sample of size n_2 drawn from a normal population with mean m_2 and unknown variance s_2^2. Then, we wish to test-

$H_0: \mu_1 = \mu_2$ (No difference between means)

$H_1: \mu_1 \neq \mu_2$ (Significance difference between means)

Test statistic

Under H_0, the test statistic used is t given by

$$t = \frac{\overline{X}-\overline{Y}}{s\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Where, $\bar{x}$ and $\bar{y}$ are the means of samples X and Y, respectively.

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2 + (\Sigma(y-\bar{y})^2}{(n_1+n_2-2)}}$$

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (n_1+n_2-2) degree of freedom. After calculating the value of the t-statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner :

i) If the calculated value of test statistic $|t| < t_{(n1+n2-2)d.f}(\alpha = 0.05)$, we accept H_0. Hence, we conclude that the two sample means do not differ significantly.

ii) If the calculated value of test statistic $|t| > t_{(n1+n2-2)d.f}(\alpha = 0.05)$, we reject H_0 at 5% level of significance for (n_1+n_2-2) degree of freedom. Hence, we conclude that the two sample means differ significantly.

iii) If the calculated value of test statistic $|t| > t_{(n1+n2-2)d.f}(\alpha = 0.01)$, we reject H_0 at 1% level of significance for (n_1+n_2-2) degree of freedom. Hence, we conclude that the two sample means **highly** differ significantly.

Example-2

Two new types of feed were fed to cows. Five cows were fed Type-A feed and another 7 cows were fed Type-B feed. The daily milk production was recorded as given below:

Type-A feed	:	10	12	13	11	14		
Type-B feed	:	8	9	12	14	15	10	9

Test whether the effect of two feeds differed significantly?

Solution:

Step-1: H_0: There is no significance difference in the efficacy of two feeds.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$t = \frac{\overline{X} - \overline{Y}}{s\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Step-4: For using test statistic, prepare the following table

X	Y	$X-\bar{X}$	$(X-\bar{X})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
10	8	-2	4	-3	9
12	9	0	0	-2	4
13	12	1	1	1	1
11	14	-1	1	3	9
14	15	2	4	4	16
	10			-1	1
9			-2	4	
$\sum X = 60$	$\sum Y = 77$		$(X-\bar{X})^2 = 10$		$(Y-\bar{Y})^2 = 44$

$$\bar{X} = \frac{\sum X}{n_1} = \frac{60}{5} = 12 \text{ and } \bar{Y} = \frac{\sum Y}{n_2} = \frac{77}{7} = 11$$

Step-5: Calculate pooled standard deviation of samples,

$$s = \sqrt{\frac{\sum(x-\bar{x})^2 + (y-\bar{y})^2}{(n_1+n_2-2)}} = \sqrt{\frac{10+44}{(5+7-2)}} = \sqrt{\frac{54}{10}} = 2.32$$

Now, the test statistic

$$t = \frac{X-\overline{X}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{12-11}{2.32\sqrt{\left(\frac{1}{5}+\frac{1}{7}\right)}} = \frac{1}{2.32\sqrt{\left(\frac{12}{35}\right)}} = \frac{1}{2.32\times 0.59} = 0.73$$

Step-6: Pooled degree of freedom = $n_1 + n_2 - 2 = 5 + 7 - 2 = 10$

Step-7: Table value of t (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$t_{10}(0.05) = 2.23 \text{ and } t_{10}(0.01) = 3.17$$

Step-8: Here, $|t_{cal}| = 0.73 < t_{tab} = 2.23$, we accept H_0.

Step-9: There is no significance difference in the efficacy of two feeds.

10.3.3 Testing the significance of the difference between two means (Paired t-test)

In the above case, we assumed that the samples have been randomly drawn from two normal populations and they are independent. However, in this situation, where two samples drawn are not independent, we use

paired t-test. Here the paired observations are recorded on the same individuals or items. Therefore, the two samples will also be of the same size n in view of the paired character or observations and may be put down as-

Let ($x_1 y_1$, (x_2,y_2), ..., (x_n,y_n) be the n pairs of observations drawn from a normal population. Let $d_i = x_i - y_i$ represent the difference for each pair (mean difference). Then, we wish to test-

H_0: $\mu_d = 0$ (Mean difference is zero)

H_1: $\mu_d \neq 0$

Test statistic

Under H_0, the test statistic used is t given by

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

Here,

$$s = \sqrt{\frac{\sum(d-\bar{d})^2}{(n-1)}}$$

Where, d = x - y

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (n-1) degree of freedom. After calculating the value of the t-statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner:

i) If the calculated value of test statistic $|t| < t_{(n-1)d.f}(\alpha = 0.05)$, we accept H_0. Hence, we conclude that the two s0ample means do not differ significantly.

ii) If the calculated value of test statistic $|t| > t_{(n-1)d.f}(\alpha = 0.05)$, we reject H_0 at 5% level of significance for (n-1) degree of freedom. Hence, we conclude that the two sample means differ significantly.

iii) If the Calculated value of test statistic it $|> t_{(n-1)d.f}(\alpha = 0.01)$, we reject H_0 at 1% level of significance for (n-1) degree of freedom. Hence, we conclude that the two sample means **highly** differ significantly.

Example-3:

Five persons participated in an experiment to study the effectiveness of a certain diet, combined with a programme of exercise in reducing serum cholesterol levels. The results are given below:

Serum cholesterol before diet	:	201	231	221	185	178
Serum cholesterol after diet	:	190	223	226	182	170

Test whether the diet exercise programme effective on the reduction in cholesterol level?

Solution:

Step-1: H_0: There is no significantly affect diet and exercise programme on the reduction in cholesterol level.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

Step-4: For using test statistic, prepare the following table

Before diet (X)	After diet (Y)	d = X - Y	$(d - \bar{d})$	$(d - \bar{d})^2$
201	190	11	6	36
231	223	8	3	9
221	226	-5	0	0
185	182	3	-2	4
178	170	8	3	9
		$\Sigma d = 25$		$\Sigma (d-\bar{d})^2 = 58$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{25}{5} = 5$$

Step-5: Calculate standard deviation of the difference,

$$s = \sqrt{\frac{\Sigma(d-\bar{d})^2}{(n-1)}} = \sqrt{\frac{58}{(5-1)}} = \sqrt{\frac{58}{4}} = 3.81$$

Now, the test statistic

$$t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{5}{3.81/\sqrt{5}} = \frac{5}{3.81/\sqrt{5}} = 2.93$$

Step-6: Degree of freedom = n – 1 = 5 – 1 = 4

Step-7: Table value of t (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$t_4(0.05) = 2.78 \text{ and } t_4(0.01) = 4.60$$

Step-8: Here, $|t_{cal}| = 2.93 > t_{tab} = 2.78$, we reject H_0 at 5% level of significance.

Step-9: There is significantly affect diet and exercise programme on the reduction in cholesterol level.

10.3.4 Testing the significance of an observed sample correlation coefficient

Let (x_1,y_1), (x_2,y_2), ..., (x_n,y_n) be the n pairs of observations drawn from a bivariate normal population. Then, we wish to test-

H_0: $\rho = 0$ (Sample correlation coefficient is zero)

H_1: $\rho \neq 0$

Test statistic

Under H_0, the test statistic used is t given by

$$t = \frac{r}{SE(r)}$$

Where,

$$SE(r) = \sqrt{\frac{(1-r^2)}{(n-2)}}$$

Here, r = correlation coefficient

and, n = number of pairs

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (n-2) degree of freedom. After calculating the value of the t-statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner:

i) If the calculated value of test statistic $|t| < t_{(n-2)d.f}(\alpha = 0.05)$, we accept H_0. Hence, we conclude that the correlation coefficient do not differ significantly.

ii) If the calculated value of test statistic $|t| > t_{(n-2)d.f}(\alpha = 0.05)$, we reject H_0 at 5% level of significance for (n-2) degree of freedom. Hence, we conclude that the correlation coefficient differ significantly.

iii) If the calculated value of test statistic $|t| > t_{(n-2)d.f}(\alpha = 0.01)$, we reject H_0 at 1% level of significance for (n-2) degree of freedom. Hence, we conclude that the correlation coefficient **highly** differ significantly.

Example-4:

From the data on age (years) and blood pressure (mm of Hg) of six adult males are given as follows:

Age (years)	:	32	48	38	24	44	34
Blood pressure (mm of Hg)	:	118	128	124	110	124	117

Test the significance of coefficient of correlation between age and blood pressure?

Solution:

Step-1: H_0: There is no significant correlation between age and blood pressure.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$t = \frac{r}{SE(r)}$$

Step-4: For using test statistic, prepare the following table

Age(X)	B.P. (Y)	$X - \bar{X}$	$(X - \bar{X})^2$	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
32	118	-5	25	-2	4	10
48	128	11	121	8	64	88
38	124	1	1	4	16	4
26	110	-11	121	-10	100	110
44	124	7	49	4	16	28
34	116	-3	9	-4	16	12
$\sum X = 222$	$\sum Y = 720$		$(X - \bar{X})^2 = 326$		$\sum (Y - \bar{Y})^2 = 216$	$\sum (X - \bar{X})(Y - \bar{Y}) = 252$

Step-5: Calculate arithmetic mean (A.M.) = $\sum X/n = 222/6 = 37$

Step-6: Calculate arithmetic mean (A.M.) = $\sum Y/n = 720/6 = 120$

Step-7: Calculate correlation coefficient between age and blood pressure, using the formula :

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} = \frac{252}{\sqrt{326\times216}} = \frac{252}{\sqrt{70416}} = \frac{252}{265.3601} = 0.9497$$

Now, the test statistic

$$t = \frac{r}{SE(r)}$$

Where,

$$SE(r) = \sqrt{\frac{(1-r^2)}{(n-2)}} = \sqrt{\frac{(1-0.9019)}{(6-2)}} = \sqrt{\frac{0.0981}{4}} = 0.1566$$

Therefore,

$$t = \frac{r}{SE(r)} = \frac{0.9497}{0.1566} = 6.07$$

Step-8: Degree of freedom = n – 2 = 6 – 2 = 4

Step-9: Table value of t (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$t_4(0.05) = 2.78 \text{ and } t_4(0.01) = 4.60$$

Step-10: Here, $|t_{cal}| = 6.07 > t_{tab} = 4.60$, we reject H_0 at 1% level of significance.

Step-11: Hence, there is highly significant correlation between age and blood pressure.

10.3.5 Testing the significance of an observed sample regression coefficient

Let (x_1,y_1), (x_2,y_2), ..., (x_n,y_n) be the n pairs of observations drawn from a bivariate normal population. Then, we wish to test-

$H_0: \beta = 0$ (Sample regression coefficient is zero)

$H_1: \beta \neq 0$

Test statistic

Under H_0, the test statistic used is t given by :

i) For the regression coefficient of y on x

$$t = \frac{b_{yx}}{SE(b_{yx})}$$

Where,

$$SE(b_{xy}) = \sqrt{\frac{\Sigma(y-\bar{y})^2 - b_{yx}\Sigma(x-\bar{x})(y-\bar{y})}{(n-2)\Sigma(x-\bar{x})^2}}$$

Here, b_{yx} = regression coefficient of y on x

and, n = number of pairs

ii) For the regression coefficient of x on y

$$t = \frac{b_{xy}}{SE(b_{xy})}$$

Where,

$$SE(b_{xy}) = \sqrt{\frac{\sum(x-\bar{x})^2 - b_{xy}\sum(x-\bar{x})(y-\bar{y})}{(n-2)\sum(y-\bar{y})^2}}$$

Here, b_{xy} = regression coefficient of x on y

and, n = number of pairs

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (n-2) degree of freedom. After calculating the value of the t-statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner-

i) If the calculated value of test statistic $|t| < t_{(n-2)d.f}(\alpha = 0.05)$, we accept H_0. Hence, we conclude that the regression coefficient do not differ significantly.

ii) If the calculated value of test statistic $|t| > t_{(n-2)d.f}(\alpha = 0.05)$, we reject H_0 at 5% level of significance for (n-2) degree of freedom. Hence, we conclude that the regression coefficient differ significantly.

iii) If the calculated value of test statistic $|t| > t_{(n-2)d.f}(\alpha = 0.01)$, we reject H_0 at 1% level of significance for (n-2) degree of freedom. Hence, we conclude that the regression coefficient **highly** differ significantly.

Example-5:

From the data on age (years) and blood pressure (mm of Hg) of six adult males are given as follows:

Age (years)	:	32	48	38	24	44	34
Blood pressure (mm of Hg)	:	118	128	124	110	124	117

Test the significance of regression of blood pressure on age?

Solution:

Step-1: H_0: There is no significant regression of blood pressure on age.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$t = \frac{b_{yx}}{SE(b_{yx})}$$

Step-4: For using test statistic, prepare the following table

Age(X)	**B.P. (Y)**	**$(x - \bar{x})$**	**$(X - \bar{x})^2$**	**$(y - \bar{y})$**	**$(Y - \bar{y})^2$**	**$(x - \bar{x})(y - \bar{y})$**
32	118	-5	25	-2	4	10
48	128	11	121	8	64	88
38	124	1	1	4	16	4
26	110	-11	121	-10	100	110
44	124	7	49	4	16	28
34	116	-3	9	-4	16	12
$\sum X = 222$	**$\sum Y = 720$**		$\sum(x-\bar{x})^2$ **= 326**		**$\sum (Y - \bar{y})^2$ = 216**	**$\sum (X - \bar{x})(Y - \bar{y})$ =252**

Step-5: Calculate arithmetic mean (A.M.) = $\sum X/n = 222/6 = 37$

Step-6: Calculate arithmetic mean (A.M.) = $\sum Y/n = 720/6 = 120$

Step-7: Calculate regression coefficient of blood pressure on age, using the formula:

$$b_{yx} = \frac{Cov(x,y)}{Var(x)} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{252}{326} = 0.7730$$

Now, the test statistic

$$t = \frac{b_{yx}}{SE(b_{yx})}$$

Where,

$$SE(b_{yx}) = \sqrt{\frac{\sum(y-\bar{y})^2 - b_{yx}\sum(x-\bar{x})(y-\bar{y})}{(n-2)\sum(x-\bar{x})^2}} = \sqrt{\frac{216-0.7730\times 252}{(6-2)\times 326}} = \sqrt{\frac{21.204}{1304}} = 0.1275$$

Therefore,

$$t = \frac{b_{yx}}{SE(b_{yx})} = \frac{0.7730}{0.1275} = 6.06$$

Step-8: Degree of freedom = n – 2 = 6 – 2 = 4

Step-9: Table value of t (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$t_4(0.05) = 2.78 \text{ and } t_4(0.01) = 4.60$$

Step-10: Here, $|t_{cal}|$ = 6.06 > t_{tab} =4.60, we reject H_0 at 1% level of significance.

Step-11: Hence, there is **highly** significant regression of blood pressure on age.

❑❑❑

Chapter 11

Chi-square Test

11.1 Introduction

Chi-square test is a **non-parametric test**. Such non-parametric tests have assumed great importance in statistical analysis and statistical inference because they are easy to compute and can be used without making assumptions about parameters as they are distribution-free tests.

11.2 Chi-square Distribution

It is a test which describes the magnitude of difference between observed frequencies and the frequencies expected under certain assumptions. The chi-square test was first used by **Karl Pearson** in the year 1900. The Chi-square test is denoted by the Greek letter 'χ^2'. Let $O_1, O_2, \ldots\ldots, O_n$ be a set of observed frequencies in classes 1, 2, ..., n and $E_1, E_2, \ldots\ldots, E_n$ be the corresponding expected frequencies, then a chi-square statistic for testing the agreement between the observed and expected frequencies is defined as :

$$\chi^2 = \sum \frac{(O-E)^2}{E} \sim \chi^2_{(n-1)d.f}$$

11.3 Application of c²-Distribution

The χ^2-distribution has a number of applications of which we will discuss the following:

i) Testing the significance for population variance

ii) Testing the goodness of fit

iii) Testing the independence of attributes

iv) Testing the independence of two attribute in a contingency table

11.3.1 Testing the significance for population variance

Let a random sample of size n drawn from a normal population with mean m and variance s^2, m and s^2 being unknown. Using the sample information, we wish to test

H_0: $\sigma^2 = \sigma_0^2$

H_1: $\sigma^2 \neq \sigma_0^2$

Where, σ_0^2 is specified value of population variance.

Test statistic

Under H_0, the test statistic used is χ^2 given by :

$$\chi^2 = \frac{\sum(x-\bar{x})^2}{\sigma_0^2} \sim \chi^2_{(n-1)d.f.}$$

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (n-1) degree of freedom. After calculating the value of the c^2-statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner:

i) If the calculated value of test statistic $\chi^2 < \chi^2_{(n-1)d.f}(\alpha = 0.05)$, we accept H_0. Hence, we conclude that the sample belongs to the population with specified variance s_0^2.

ii) If the calculated value of test statistic $\chi^2 > \chi^2_{(n-1)d.f}(\alpha = 0.05)$, we reject H_0 at 5% level of significance for (n-1) degree of freedom. Hence, we conclude that the sample does not belong to the population with specified variance s_0^2.

iii) If the calculated value of test statistic $\chi^2 > \chi^2_{(n-1)d.f}(\alpha = 0.01)$, we reject H_0 at 1% level of significance for (n-1) degree of freedom. Hence, we conclude that the sample highly different to the population with specified variance σ^2.

Example-1:

Body weights in kg of 5 students are given below:

53	55	48	52	62

Test whether the distribution of body weights of all students from which the above sample belong to the population having variance 20 square kg?

Solution:

Step-1: H_0: The distribution of sample belongs to the population having variance 20 square kg.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$\chi^2 = \frac{\sum(x-\bar{x})^2}{\sigma_0^2} \sim \chi^2_{(n-1)d.f.}$$

Step-4: For using test statistic, prepare the following table

Body weight (X)	$(x - \bar{x})$	$(x - \bar{x})^2$
53	-1	1
55	1	1
48	-6	36
52	-2	4
62	8	64
$\sum$ X =270		$\sum (X - \bar{X})^2 = 106$

Step-5: Calculate arithmetic mean (A.M.) = $\sum X/n = 270/5 = 54$

Step-6: Now the test statistic:

$$\chi^2 = \frac{\sum(x-\bar{x})^2}{\sigma_0^2} = \frac{106}{20} = 5.3$$

Step-7: Degree of freedom = n – 1 = 5 – 1 = 4

Step-8: Table value of χ^2 (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$\chi^2(0.05) = 11.07 \text{ and } \chi_4^2(0.01) = 15.09$$

Step-9: Here, χ^2_{cal} = 5.3 < =11.07, we accept H_0.

Step-10: Hence, the distribution of sample belongs to the population having variance 20 square kg.

11.3.2 Testing the goodness of fit

The test for goodness of fit determines whether a population has a specified theoretical distribution. In other words, here our problem is to test the hypothesis of how closely the observed distribution approximates a particular theoretical distribution. Using the sample information, we wish to test

H_0: O = E (Provide good fit)

H_1: O $\neq$ E

Where, O = Observed frequencies

E = Expected frequencies

Test statistic

Under H_0, the test statistic used is c^2 given by :

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} \sim \chi^2_{(n-1)d.f}$$

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (n-1) degree of freedom. After calculating the value of the c^2-statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner:

i) If the calculated value of test statistic $\chi^2 < \chi^2_{(n-1)d.f}$ (α = 0.05), we accept H_0. Hence, we conclude that the observed data follow the theory.

ii) If the calculated value of test statistic $\chi^2 > \chi^2_{(n-1)d.f}$ (α = 0.05), we reject H_0 at 5% level of significance for (n-1) degree of freedom. Hence, we conclude that the observed data not follow the theory.

iii) If the calculated value of test statistic $\chi^2 > \chi^2_{(n-1)d.f}$ (α = 0.01), we reject H_0 at 1% level of significance for (n-1) degree of freedom. Hence, we conclude that the observed data highly different to the theory.

Example-2:

The number of theory classes held during different months in a particular semester of biostatistics in PG programme is given below:

Month	:	First	Second	Third	Fourth	Fifth
No. of classes held	:	12	10	9	14	5

Test whether the classes are normally distributed over the semester?

Solution:

Step-1: H_0: The classes are normally distributed over the semester.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} \sim \chi^2_{(n-1)d.f}$$

Step-4: For using test statistic, prepare the following table

Observed (O)	O – E	(O – E)^2
53	-1	1
55	1	1
48	-6	36
52	-2	4
62	8	64
Σ **O =270**		$\Sigma(O-E)^2$ **= 106**

Step-5: Calculate the expected frequency of classes will be E = Σ O/ n = 270/5 = 54

Step-6: Now the test statistic :

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{\sum (O-E)^2}{E} = \frac{106}{54} = 1.962$$

Step-7: Degree of freedom = n – 1 = 5 – 1 = 4

Step-8: Table value of [redacted] (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$\chi^2_4(0.05) = 11.07 \text{ and } \chi^2_4(0.01) = 15.09$$

Step-9: Here, $_{cal}$ = 1.962 < 11.07, we accept H_0.

Step-10: Hence, the classes are normally distributed over the semester.

11.3.3 Testing the independence of attributes

In this case, there are no definite expected values, the point is whether the results are dependent or independent of the conditions under which they have occurred. This test is called a **test for independence or contingency test**. The observed frequencies are indicated in various cells of the table for respective rows and columns. If there are 'm' rows and 'n' columns, the table is generally called "m x n" **contingency table**.

Suppose N observations in a sample are to be classified according to two attributes say A and B. Attribute A has m mutually exclusive categories say A_1, A_2, ………., A_m and the attribute B has categories namely B_1, B_2, …………., B_n. Then the sample observations may be classified as :

Contingency table of order m x n

B **A**	**B_1**	**B_2**	--------	**B_j**	--------	**B_n**	**Total**
A_1	O_{11}	O_{12}	--------	O_{1j}	--------	O_{1n}	**(A_1)**
A_2	O_{21}	O_{22}	--------	O_{2j}	--------	O_{2n}	**(A_2)**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_i	O_{i1}	O_{i2}	--------	O_{ij}	--------	O_{in}	**(A_i)**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_m	O_{m1}	O_{m2}	--------	O_{mj}	--------	O_{mn}	**(A_m)**
Total	**(B_1)**	**(B_2)**	--------	**(B_j)**	--------	**(B_n)**	**N**

The above two-way table having m rows and n columns is called a **contingency table** of order (m x n). In this table-

A_i denotes the i^{th} category of the attribute A. (i = 1, 2, …….., m)

B_j denotes the j^{th} category of the attribute A. (j = 1, 2, …….., n)

(A_i) denotes the frequency of the attribute A_i.

(B_j) denotes the frequency of the attribute B_j.

$\sum_{i=1}^{m}(A_i) = \sum_{j=1}^{n}(B_j) = N$, the total number of observations.

(O_{ij}) is the observed frequency in (I, j) cell.

Suppose we are given a contingency table of order m x n in which N sample observations have been classified. Let O_{ij} be the observed frequency in $(i, j)^{th}$ cell. To test the null hypothesis that the two attributes are independent, we use chi-square test.

Using the sample information, we wish to test

H_0: Two attributes A and B are independent

H_1: Two attributes A and B are dependent.

Test statistic

Under H_0, the test statistic used is given by :

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \sim \chi^2_{(m-1)(n-1)d.f.}$$

Here, the χ^2 - distribution has (m-1)(n-1) degree of freedom. The expected frequency corresponding to (I, j)th cell observed frequency i.e. E_{ij} is obtained by :

$$E_{ij} = E(O_{ij}) = \frac{(A_i)(B_j)}{N} = \frac{Sum\ of\ i^{th}\ row \times Sum\ of\ j^{th}\ column}{Sample\ size}$$

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (m-1)(n-1) degree of freedom. After calculating the value of the c^2-statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner :

i) If the calculated value of test statistic $\chi^2 < \chi^2_{(m-1)(n-1)d.f}(\alpha = 0.05)$, we accept H_0. Hence, we conclude that the two attributes are independent.

ii) If the calculated value of test statistic $\chi^2 > \chi^2_{(m-1)(n-1)d.f}(\alpha = 0.05)$, we reject H_0 at 5% level of significance for (m-1)(n-1) degree of freedom. Hence, we conclude that the two attributes are dependent or there is a significant relationship between the two variables.

iii) If the calculated value of test statistic $\chi^2 > \chi^2_{(m-1)(n-1)d.f}(\alpha = 0.01)$, we reject H_0 at 1% level of significance for (m-1)(n-1) degree of freedom. Hence, we conclude that there is a highly significant relationship between the two variables.

Example-3:

Patients of 200 heart diseases were classified according to age and type of heart disease. The results are given below:

Age	Coronary	Heart disease		Other	Total
		Coronary and hypertensive	Hypertensive		
20-50	27	16	29	18	**90**
51-70	29	24	35	22	**110**
Total	**56**	**40**	**64**	**40**	**200**

Is there any relationship between age and type of heart disease?

Solution:

Step-1: H_0: There is no relationship between age and type of heart disease.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(m-1)(n-1)\text{d.f.}}$$

Step-4: Calculate the expected frequencies corresponding to each observed frequency as follows:

$$E_{11} = \frac{90\times 56}{200} = 25.2; E_{12} = \frac{90\times 40}{200} = 18; E_{13} = \frac{90\times 64}{200} = 28.8$$

E_{21} = 56 – 25.2 = 30.8; E_{22} = 40 – 18 = 22; E_{23} = 64 – 28.8 = 35.2; E_{14} = 90 – (25.2 + 18 + 28.8) = 18

E_{24} = 40 – 18 = 22

Step-5: For using test statistic, prepare the following table of observed and expected frequency:

Observed (O)	Expected (E)	O – E	(O – E)^2	(O – E^2/E
27	25.2	1.8	3.24	0.1286
16	18	-2	4	0.2222
29	28.8	0.2	0.04	0.0014
18	18	0	0	0
29	30.8	-1.8	3.24	0.1052
24	22	2	4	0.1818
35	35.2	- 0.2	0.04	0.0011
22	22	0	0	0
Σ O =200	**Σ E =200**			**Σ (O – E)^2/E =0.6403**

Step-6: Now the test statistic :

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = 0.6403$$

Step-7: Here, degree of freedom = (r – 1)(c – 1) = (2 – 1)(4 – 1) = 3

Step-8: Table value of  (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$\chi_3^2\,(0.05) = 7.81\, and\ \chi_3^2\,(0.01) = 11.34$$

Step-9: Here, $\chi^2_{cal} = 0.6403 < 7.81$, we accept H_0.

Step-10: Hence, there is no relationship between age and type of heart disease.

11.3.4 Testing the independence of attributes in a contingency table

Sometimes the data are cross-classified in such a manner that there are only two categories. The contingency table containing such data which consist of two rows and two columns is often referred to as 2 x 2 table. Then the sample observations may be classified as :

Contingency table of order m x n

B/A	**B_1**	**B_2**	**Total**
A_1	O_{11}(a)	O_{12} (b)	**(A_1)**
A_2	O_{21} (c)	O_{22} (d)	**(A_2)**
Total	**(B_1)**	**(B_2)**	**N**

The above two-way table having 2 rows and 2 columns is called a **contingency table** of order (2 x 2). Using the sample information, we wish to test :

H_0: Two attributes A and B are independent.

H_1: Two attributes A and B are dependent.

Test statistic

Under H_0, the test statistic used is given by :

$$\chi^2 = \frac{(ad-bc)^2 \times N}{(a+b)(c+d)(a+c)(b+d)}$$

Here, the c^2 - distribution has (2-1)(2-1) = 1 x 1 =1 degree of freedom.

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for **one** degree of freedom. After calculating the value of the χ^2 -statistic, the decision about the acceptance or rejection of H_0 is taken in the following manner :

i) If the calculated value of test statistic $\chi^2 < \chi^2_{1d.f}(\alpha = 0.05)$, we accept H_0. Hence, we conclude that the two attributes are independent.

ii) If the calculated value of test statistic $\chi^2 > \chi^2_{1d.f}(\alpha = 0.05)$, we reject H_0 at 5% level of significance for 1 degree of freedom. Hence, we conclude that the two attributes are dependent or there is a significant relationship between the two variables.

iii) If the calculated value of test statistic $\chi^2 > \chi^2_{1d.f}(\alpha = 0.01)$, we reject H_0 at 1% level of significance for 1 degree of freedom. Hence, we conclude that there is a highly significant relationship between the two variables.

Example-4:

Following table shows the results of a particular drug against heart attacks:

Group	Not attacked	Attacked	Total
Drug given	12	8	**20**
No drug	16	24	**40**
Total	**28**	**32**	**60**

Find out whether there is any significant association between drug and attack.

Solution:

Step-1: H_0: There is association between drug and attack.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$\chi^2 = \frac{(ab-bc)^2 \times N}{(a+b)(c+d)(a+c)(b+d)} = \frac{(12\times 24 - 16\times 8)^2 \times 60}{28\times 32\times 20\times 40} = \frac{(160)^2 \times 60}{716800} = 2.14$$

Step-4: Here, degree of freedom = (r – 1)(c – 1) = (2 – 1)(2 – 1) = 1

Step-5: Table value of χ^2 (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$\chi^2_1(0.05) = 3.84 \text{ and } \chi^2_1(0.01) = 6.64$$

Step-6: Here, $\chi^2_{cal} = 2.14 < 3.84$, we accept H_0.

Step-7: Hence, there is association between drug and attack.

Yate's correction for continuity

If any of the observed cell frequency is less than 5, in a 2 x 2 contingency table. Yate's correction for continuity may be used. In this case, by utilizing the following formula :

$$\chi^2 = \frac{\left(|ad - bc| - \frac{N}{2}\right)^2 \times N}{(a+b)(c+d)(a+c)(b+d)}$$

It should be noted that this correction should not be used when m > 2 and n > 2.

Example-5:

In an experiment on the immunization of goats from a disease, the following results were obtained:

Group	Survived	Died of disease	Total
Inoculated	16	4	**20**
Not inoculated	18	12	**30**
Total	**34**	**16**	**50**

Test the effect of inoculation on the control of disease.

Solution:

Step-1: H_0: Inoculation and disease are independent.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: For testing the hypothesis, we use test statistic

$$\chi^2 = \frac{\left(|ad-bc| - \frac{N}{2}\right)^2 \times N}{(a+b)(c+d)(a+c)(b+d)} = \frac{\left(|16\times12-18\times4| - \frac{50}{2}\right)^2 \times 50}{326400} \quad \frac{\left(|120| - \frac{50}{2}\right)^2 \times 50}{} = 1.38$$

Step-4: Here, degree of freedom = (r – 1)(c – 1) = (2 – 1)(2 – 1) = 1

Step-5: Table value of [redacted] (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$\chi_1^2(0.05) = 3.84 \; and \; \chi_1^2(0.01) = 6.64$$

Step-6: Here, χ^2_{cal} = 1.38 < 3.84, we accept H_0.

Step-7: Hence, Inoculation and disease are independent.

❑❑❑

Chapter 12

F - Test

12.1 Introduction

The name was coined by **George W. Snedecor**, in honour of **Sir Ronald A. Fisher**. The object of the F-test is to find out whether the two independent estimates of population variance differ significantly or whether the two samples may be regarded as drawn from the normal populations having the same variance. It is defined as :

$$F = \frac{s_1^2}{s_2^2} \sim F_{(n_1-1)(n_2-1)d.f}$$

Where, $s_1^2 > s_2^2$ and

$$s_1^2 = \frac{\Sigma(x-\bar{x})^2}{(n_1-1)}$$

$$s_1^2 = \frac{\Sigma(y-\bar{y})^2}{(n_2-1)}$$

Here, the F-test has two degree of freedoms i.e. $(n_1\text{-}1)$ degree of freedom for numerator and $(n_2\text{-}1)$ degree of freedom for denominator.

12.2 Application of F-test

The F-test has a number of applications of which we will discuss the following:

i) Testing the significance of ratio of two variance

ii) Testing the homogeneity of several means.

12.2.1 Testing the significance of ratio of two variance

Suppose x_1, x_2, …………., x_n be a random sample of size n_1 drawn from a normal population with mean m_1 and variance s_1^2 and another random sample of size n_2 drawn from a normal population with mean m_2 and variance s_2^2. The null and alternative hypothesis are :

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

Test statistic

Under H_0, the test statistic F is given by :

$$F = \frac{S_1^2}{S_2^2}$$

Where, $S_1^2 > S_2^2$ and

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{(n_1-1)}$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{(n_2-1)}$$

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for $(n_1$-1),$(n_2$-1) degree of freedom. After calculating the value of the F-test, the decision about the acceptance or rejection of H_0 is taken in the following manner :

i) If the calculated value of test statistic $F \leq F_{(n_1-1),(n_2-1)d.f}$ (α = 0.05), we accept H_0. Hence, we conclude that the ratio of two variances do not differ significantly.

ii) If the calculated value of test statistic $F > F_{(n_1-1),(n_2-1)d.f}$ (α = 0.05), we reject H_0 at 5% level of significance for $(n_1$-1),$(n_2$-1) degree of freedom. Hence, we conclude that the ratio of two variances differ significantly.

iii) If the calculated value of test statistic $F > F_{(n_1-1),(n_2-1)d.f}$ (α = 0.01), we reject H_0 at 1% level of significance for $(n_1$-1),$(n_2$-1) degree of freedom. Hence, we conclude that the ratio of two variances **highly** differ significantly.

12.2.2 Testing the homogeneity of several means

In this case we should have a technique by which significance of the difference amongst more than two sample means is carried out at the same time. The **analysis of variance (ANOVA)** technique enables us to perform this simultaneous test.

12.3 Analysis of Variance

The analysis of variance technique, developed by **R.A. Fisher** in 1920, is capable of fruitful application to a diversity of practical problems. The '**Analysis of Variance**' is a technique of partitioning or splitting the total variation i.e. assignable factors and the chance factors and then their comparison. The analysis of variance frequently referred to by the contraction ANOVA is a statistical technique specially designed to test whether the means of more than two sample means is carried out at the same time.

12.3.1 Classification of observations

The following criteria of classification of observations are used in the analysis of variance.

i) One-way classification

ii) Two-way classification

One-way classification

In this case the observations are classified on the basis of a single criterion, so the classification is called **one-way classification**.

Two-way classification

If the observations in an experiment are classified in respect of two criteria, we get **two-way classification**.

12.3.1.1 One-way Classification of Analysis of Variance

Suppose there are k normal populations with means m_1, m_2, ……., m_k and common variance s^2. Further, let k random samples, one from each population, are drawn from these populations. Let n_i (i = 1, 2, …..,k) be the size of the sample from i^{th} population. Using the sample information, we wish to test-

H_0: $\mu_1 = \mu_2 = \ldots\ldots\ldots\ldots = \mu_k$

H_1: $\mu_1 \neq \mu_2 \neq \ldots\ldots\ldots\ldots \neq \mu_k$

Let y_{ij} (i = 1, 2, ………., k; j = 1, 2, …………., n_i) be the j^{th} observation of i^{th} sample, then one-way classified data can be arranged as :

One way classification

Sample No.	Observations							Sample Totals	Sample Means
1	Y_{11}	Y_{12}	Y_{13}	...	...	...	y_{1n_1}	$T_{1.}$	$\bar{y}_{1.}$
2	Y_{21}	Y_{22}	Y_{23}	...	...	...	y_{2n_2}	$T_{2.}$	$\bar{y}_{2.}$
3	Y_{31}	Y_{32}	Y_{33}	...	...	...	y_{3n_3}	$T_{3.}$	$\bar{y}_{3.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	y_{k_1}	y_{k_2}	y_{k_3}	...	...	...	y_{kn_k}	$T_{k.}$	$\bar{y}_{k.}$
Over all total								$T_{..}$	$\bar{y}_{..}$

Here,

$$T = \sum_{i=1}^{k} T_i$$

and

$$\bar{y} = \sum_{i=1}^{k} \bar{y}_i$$

also let

$$N = \sum_{i=1}^{k} n_i$$

Now, for testing H_0, the steps in the analysis of variance procedure for one-way classification are-

i) Calculate the correction factor (C.F.) as

$$C.F. = \frac{T^2}{N}, N = \sum_{i=1}^{k} n_1 = (n_1 + n_2 + ... + n_k)$$

ii) Find to total sum of squares (TSS)

$$TSS = \sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_i \sum_j y_{ij}^2 - C.F. = (y_{11}^2 + y_{12}^2 + ... + y_{1n_1}^2 + ... + y_{kn_k}^2) - C.F.$$

iii) Find the between sample sum of squares (BSS)

$$BSS = \sum_i \sum_j (\bar{y}_i - \bar{y})^2 \sum_{i=1}^{n} \frac{T_1^2}{n} - C.F$$

iv) Find error sum of squares (ESS) by subtraction. i.e. ESS = TSS - BSS

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2=\sum_{i=1}^{k}\sum_{j=1}^{n_j}(y_{ij}-\bar{y})^2-\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_i-\bar{y})^2$$

v) Prepare the ANOVA (Analysis of variance) table

ANOVA -table (One-way classification)

Sources of variations	Degree of freedom (d.f.)	Sum of Squares (SS)	Mean Sum of Squares (MSS)	Calculated variance ration (F)
Between Samples	k - 1	$\sum_i\sum_j(\bar{y}_{i.}-\bar{y}_{..})^2$	$\frac{\sum_i\sum_j(\bar{y}_{i.}-\bar{y}_{..})^2}{k-1}=V_1(\text{say})$	$\frac{V_1}{V_2}=F$
Within Samples or Error	N - k	$\sum_i\sum_j(\bar{y}_{ij}-\bar{y}_{i.})^2$	$\frac{\sum_i\sum_j(\bar{y}_{ij}-\bar{y}_{i.})^2}{N-k}=V_2(\text{say})$	
Total	N - 1	$\sum_i\sum_j(y_{ij}-\bar{y}_{..})^2$		

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (k-1),(N-k) degree of freedom. After calculating the value of the F-test, the decision about the acceptance or rejection of H_0 is taken in the following manner :

i) If the calculated value of test statistic $F \le F_{(k-1),(N-k)d.f}$ (α = 0.05), we accept H_0. Hence, we conclude that the samples do not differ significantly.

ii) If the calculated value of test statistic $F > F_{(k-1),(N-k)d.f}$ (α = 0.05), we reject H_0 at 5% level of significance for (k-1),(N-k) degree of freedom. Hence, we conclude that the samples differ significantly.

iii) If the calculated value of test statistic $F > F_{(k-1),(N-k)d.f}$ (α = 0.01), we reject H_0 at 1% level of significance for (k-1),(N-k) degree of freedom. Hence, we conclude that the samples **highly** differ significantly.

Example-1:

Compare the performance of four different hospitals with regard to the number of heart disease cases successfully treated per month. A sample of ten records of each hospital and the number of heart diseases cases successfully treated was given below:

						Total	Average
Hospital-A	7	6	8	7	6	34	6.4
Hospital-B	6	7	5	8	7	33	6.6
Hospital-C	7	8	9	5	6	35	7.0
Hospital-D	7	6	8	7	6	34	8.0
Total						136	

Test whether a difference in the number of heart disease cases successfully treated per month among four hospitals.

Solution:

Step-1: H_0: There is no significant difference between hospitals regarding the number of heart disease cases successfully treated per month.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: Calculate correction factor as:

$$C.F. = \frac{T^2}{N} = \frac{(136)^2}{20} = 924.8$$

Step-4: Now, computation of various sums of squares:

Step-5: Calculate Sum of square

$$\sum_i \sum_j yij = (7^2+6^2+8^2+7^2+6^2+6^2+7^2+5^2+8^2+7^2+7^2+8^2+9^2+5^2+6^2+7^2+6^2+8^2+7^2+6^2) = 946$$

Step-6: Calculate Total Sum of Square (TSS)

$$TSS = \sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_i \sum_j y_{ij}^2 - C.F. = 946 - 924.8 = 21.20$$

Step-7: Calculate between sample (hospital) Sum of square

$$BSS = \sum_i \sum_j (\bar{y}_{ij} - \bar{y})^2 = \sum_{i=1}^{n} \frac{T_i^2}{n} - C.F. = \frac{(34^2+33^2+35^2+34^2)}{5} - 924.80 = 0.40$$

Step-8: Calculate error sum of squares (ESS):

ESS = TSS – BSS = 21.20 – 0.40 = 20.80

Step-9: Prepare the ANOVA (Analysis of variance) table:

Sources of variations	Degree of freedom (d.f.)	Sum of Squares (SS)	Mean Sum of Squares (MSS)	Calculated variance ration (F)
Between Samples (Hospitals)	k – 1 = 4-1=3	0.40	$\frac{0.40}{3} = 0.1333$	$\frac{0.1333}{1.30} = 0.1025$
Within Samples (hospitals) or Error	N – k = 20 – 4=16	20.80	$\frac{20.80}{16} = 1.30$	
Total	N – 1= 20 – 1 = 19	21.20		

Step-10: Here, degree of freedom

i) for numerator: k – 1 = 4 – 1 = 3

ii) for denominator: N – k = 20 – 4 = 16

Step-11: Table value of F (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$F_{3,16}(0.05) = 3.24$ and $F_{3,16}(0.01) = 5.29$

Step-12: Here, $F_{cal} = 0.1025 < F_{3,16}(0.05) = 3.24$, we accept H_0.

Step-13: Hence, there is no significant difference between hospitals regarding the number of heart disease cases successfully treated per month.

12.3.1.2 Two-way Classification of Analysis of Variance

Suppose there are k rows and n columns in the rectangular array representing a two-way classification. Let y_{ij} denotes a cell observation in i^{th} row and j^{th} column (i=1, 2, …….,k; j=1, 2, ……., n). Then, a two-way classification model can be represented as-

Two-way classification

Rows	Columns							Total ($T_{i.}$)	Mean ($\bar{y}_{i.}$)
	1	2	3	…	j	…	n		
1	Y_{11}	Y_{12}	Y_{13}	…	y_{ij}	…	y_{1n}	$T_{1.}$	$\bar{y}_{1.}$
2	Y_{21}	Y_{22}	Y_{23}	…	Y_{2j}	…	y_{2n}	$T_{2.}$	$\bar{y}_{2.}$
3	Y_{31}	Y_{32}	Y_{33}	…	Y_{3j}	…	y_{3n}	$T_{3.}$	$\bar{y}_{3.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	Y_{i1}	Y_{i2}	Y_{i3}	…	y_{ij}	…	y_{in}	$T_{i.}$	$\bar{y}_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	y_{k1}	y_{k2}	y_{k3}	…	y_{kj}	…	y_{kn}	$T_{k.}$	$\bar{y}_{k.}$
Total ($T_{.j}$)	$T_{.1}$	$T_{.2}$	$T_{.3}$	…	$T_{.j}$	…	$T_{.n}$	$T_{..}$	
Mean ($\bar{y}_{.j}$)	$\bar{y}_{.1}$	$\bar{y}_{.2}$	$\bar{y}_{.3}$	…	$\bar{y}_{.j}$	…	$\bar{y}_{.n}$		$\bar{y}_{..}$

Here, $T_{i.}$ and $\bar{y}_{i.}$ respectively stand for total and mean of the i^{th} row $(i = 1, 2, \ldots\ldots, k)$

$T_{.j}$ and $\bar{y}_{.j}$ respectively stand for total and mean of the j^{th} column $(j = 1, 2, \ldots\ldots, n)$ and

$T_{..}$ and $\bar{y}_{..}$ respectively stand for total and mean of all the nk observations.

Further, suppose the sample observations in i^{th} row be drawn from a normal population with mean m_{i}. and variance s^2. Similarly, the sample observations in j^{th} column are drawn from a normal population with mean $m_{.j}$ and variance s^2. Using the sample information, we wish to set up the following two null hypotheses :

a) $H_{01}: m_{1.} = m_{2.} = \ldots\ldots = m_{i.} = \ldots.. = m_{k.}$ (i.e. rows means are equal)

$H_{11}: m_{1.} \neq m_{2.} \neq \ldots\ldots \neq m_{i} \neq \ldots.. \neq m_{k}$ (i.e. rows means are not equal)

b) $H_{02}: m_{.1} = m_{.2} = \ldots\ldots = m_{.j} = \ldots.. = m_{.n}$ (i.e. column means are equal)

$H_{12}: m_{.1} \neq m_{.2} \neq \ldots\ldots \neq m_{.j} \neq \ldots.. \neq m_{.n}$ (i.e. column means are not equal)

Now, for testing H_0, the steps in the analysis of variance procedure for two-way classification are :

i) Calculate the correction factor (C.F.) as :

$$C.F. = \frac{T^2}{nk}$$

ii) Find to total sum of squares (TSS)

$$TSS = \sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_i \sum_j y_{ij}^2 - C.F. = (y_{11}^2 + y_{12}^2 + \ldots. + y_{kn}^2) - C.F$$

iii) Find the rows sum of squares (RSS)

$$RSS = \sum_i \sum_j (y_i - \bar{y})^2 = \sum_{i=1}^{k} \frac{T_i^2}{n} - C.F$$

iv) Find the column sum of squares (CSS)

$$CSS = \sum_i \sum_j (Y_j - \bar{Y})^2 = \sum_{i=1}^{k} \frac{T_j^2}{n} - C.F$$

v) Find error sum of squares (ESS) by subtraction. i.e.

ESS = TSS - (RSS + CSS)

vi) Prepare the ANOVA (Analysis of variance) table

ANOVA -table (Two-way classification)

Sources of variations	Degree of freedom (d.f.)	Sum of Squares (SS)	Mean Sum of Squares (MSS)	Calculated (F)	Tabulated (F)
Row	k - 1	RSS	$\frac{RSS}{k-1} = V_1(\text{say})$	$\frac{V_1}{V_3} = F_1$	$F_{v_1,v_3}(\alpha)$
Column	n - 1	CSS	$\frac{CSS}{n-1} = V_2(\text{say})$	$\frac{V_2}{V_3} = F_2$	$F_{v_2,v_3}(\alpha)$
Error	(k - 1)(n-1)	ESS	$\frac{ESS}{(k-1)(n-1)} = V_3(\text{say})$		
Total	nk - 1	TSS			

Here, $n_1 = (k-1)$, $n_2 = (n-1)$ and $n_3 = (k-1)(n-1)$

Decision rule

After forming the above ANOVA table, the decisions regarding H_{01} and H_{02} are taken as :

a) The test is usually performed at 5% i.e. ($\alpha = 0.05$) and 1% i.e. ($\alpha = 0.01$) level of significances for n_1,n_3 degree of freedom. After calculating the value of the F-test, the decision about the acceptance or rejection of H_{01} is taken in the following manner :

i) If the calculated value of test statistic $F \le F_{V1/V3\,d.f}$ ($\alpha = 0.05$), we accept H_{01}. Hence, we conclude that the row samples do not differ significantly.

ii) If the calculated value of test statistic $F > F_{V1/V3\,d.f}$ ($\alpha = 0.05$), we reject H_{01} at 5% level of significance for n_1,n_3 degree of freedom. Hence, we conclude that the row samples differ significantly.

iii) If the calculated value of test statistic $F \le F_{V1/V3\,d.f}$ ($\alpha = 0.01$), we reject H_{01} at 1% level of significance for n_1,n_3 degree of freedom. Hence, we conclude that the row samples **highly** differ significantly.

b) The test is usually performed at 5% i.e. ($\alpha = 0.05$) and 1% i.e. ($\alpha = 0.01$) level of significances for n_2,n_3 degree of freedom. After calculating the value of the F-test, the decision about the acceptance or rejection of H_{02} is taken in the following manner :

i) If the calculated value of test statistic $F \le F_{v_2 v_3 d.f.}(\alpha = 0.05)$ we accept H_{02}. Hence, we conclude that the column samples do not differ significantly.

ii) If the calculated value of test statistic $F > F_{v_2 v_3 d.f.}(\alpha = 0.05)$, we reject H_{02} at 5% level of significance for n_2, n_3 degree of freedom. Hence, we conclude that the column samples differ significantly.

iii) If the calculated value of test statistic $F > F_{v_2 v_3 d.f.}(\alpha = 0.01)$, we reject H_{02} at 1% level of significance for n_2, n_3 degree of freedom. Hence, we conclude that the column samples **highly** differ significantly.

12.4 Critical Difference Test (CD-test)

When values of F are found **significant** then **critical difference test (CD-test)** is applied. By this test, the different groups of assignable factors can be graded according to their magnitude i.e. superiority in some classes or groups can be established over other classes or groups.

12.4.1 Procedure of critical difference test (CD-test)

For calculating the value of critical difference, using following formula-

$$C.D. = t_{error\ d.f.(0.05)} \times \sqrt{\frac{2 \times EMSS}{r}}$$

Where,

t = table value of t-test with error degree of freedom mentioned in the ANOVA-table and 5% level of significance.

EMSS = error mean sum of square mentioned in the ANOVA-table

r = number of replications

The means of different classes of assignable factors are arranged either in ascending order or in descending order of their magnitude. Then, the actual difference between the different pairs will be compared with the value of critical difference (C.D.). If actual difference is found less than the value of C.D., it means that the two means do not differ significantly and they are connected with a **straight line**.

On the other hand, if actual difference is more or equal to the value of C.D. then the two means are said to be significantly different and they are **not** joined with the straight line. In this way, we can grade the means of different classes and the superiorities of some classes can be established over the other classes.

Example-2:

Three doctors each tests five treatment for a certain disease and observe the number of days, each patient take to recover. The following table gives the recovery times in days corresponding to each doctor and each treatment.

Doctor	Treatments					Total	Average
	I	II	III	IV	V		
A	11	15	24	20	23	93	18.6
B	12	16	17	18	20	83	16.6
C	10	14	19	16	21	80	16.0
Total	33	45	60	54	64	256	
Average	11.0	15.0	20.0	18.0	21.33		

Test whether there is any significant difference between the doctors and between the treatments.

Solution:

Step-1: H_0: There is no significant difference between the doctors and between the treatments.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: Calculate correction factor as:

$$C.F. = \frac{T^2}{nk} = \frac{(256)^2}{5x3} = 4369.067$$

Step-4: Now, computation of various sums of squares:

Step-5: Calculate Sum of square

$$\sum_i \sum_j yij = (11^2 + 12^2 + 10^2 + 15^2 + 16^2 + 14^2 + 24^2 + 17^2 + 19^2 + 20^2 + 18^2 + 16^2 + 23^2 + 20^2 + 21^2) = 4618$$

Step-6: Calculate Total Sum of Square (TSS)

$$TSS = \sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_i \sum_j y_{ij}^2 - C.F. = 4618 - 4369.067 = 248.9333$$

Step-7: Calculate row (doctors) Sum of square (RSS)

$$RSS = \sum_i \sum_j (\bar{y}_1 - \bar{y})^2 = \sum_{i=1}^{k} \frac{T_i^2}{n} - C.F. = \frac{93^2 + 83^2 + 80^2}{5} - 4369.67 = 18.5333$$

Step-8: Calculate $S_1^2 = \frac{\Sigma(x-\bar{x})^2}{(n_1-1)}$ column (treatments) Sum of square (CSS)

Step-9: Calculate error sum of squares (ESS):

ESS = TSS – (BSS + CSS) = 248.9333 – (18.5333 + 206.2667) = 24.1333

Step-10: Prepare the ANOVA (Analysis of variance) table:

Sources of variations	Degree of freedom (d.f.)	Sum of Squares (SS)	Mean Sum of Squares (MSS)	Calculated variance ration (F)
Row	k – 1 = 3-1=2	18.5333	$\frac{18.5333}{2} = 9.2667$	$\frac{9.2667}{3.0167} = 3.0718$
Column	n -1 = 5 – 1 = 4	206.2667	$\frac{206.2667}{4} = 51.5667$	$\frac{51.5667}{3.0167} = 17.0937$
Error	(n – 1)(k -1) = 8	24.1333	$\frac{24.1333}{8} = 3.0167$	
Total	nk – 1= 5x3 – 1 = 14	248.5333		

Step-11: Here, degree of freedom (for rows)

i) for numerator: k – 1 = 3 – 1 = 2

ii) for denominator: (n – 1)(k – 1) = (4 – 1)(3 – 1) = 8

Step-12: Here, degree of freedom (for columns)

i) for numerator: n – 1 = 5 – 1 = 4

ii) for denominator: (n – 1)(k – 1) = (4 – 1)(3 – 1) = 8

Step-13: Table value of F – for rows (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$F_{2,8}(0.05) = 4.46$ and $F_{2,8}(0.01) = 8.65$

Step-14: Here, F_{cal} = 3.0718 < $F_{2,8}(0.05)$ =4.46, we accept H_0.

Step-15: Table value of F – for columns (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$F_{4,9}(0.05) = 3.84$ and $F_{4,9}(0.01) = 7.01$

Step-16: Here, $_{cal}$ = 17.0937 > =7.01, we reject H_0 at 1% level of significance.

Step-17: Hence, there is no significant difference between the doctors but there is highly significant difference between the treatments.

Step-18: Now, if the significant difference between the treatments, then we apply critical difference (CD) test.

Step-19: For calculating the value of critical difference, using following formula :

$$C.D. = t_{error\,d.f.}(0.05) \times \sqrt{\frac{2 \times EMSS}{r}} = 2.306 \times \sqrt{\frac{2 \times 3.0167}{3}} = 3.27$$

Step-20: Now, means of different classes (Treatments) in ascending order:

Treatments:	I	II	IV	III	V
Means:	11.0	15.0	18.0	20.0	21.33

Step-21: Hence,

i) Treatment III and V shows the similar effect and superior than other treatments.

ii) Treatment III and IV shows the similar effect but superior than treatment I and II.

iii) Treatment II and IV shows the similar effect but superior than treatment I.

❑❑❑

Chapter 13

Design of Experiment

13.1 Introduction

In 1935 **Prof. Ronald A. Fisher** laid the foundation for the subject in his monumental work entitled **"The Designs of Experiments"**. Experimental designs concern the arranging of treatments or variables in such a manner that the inferences and conclusions regarding the effects of these treatments can be easily done and their reliability measured. Experiments are made with a view to find the validity of a particular hypothesis and to have an idea about the extent of the reliability that can be placed on a particular conclusion arrived. The selection of the design will have a very great bearing on the accuracy of the ultimate results. By a random selection of experimental units it is possible to remove the ambiguity about the casual interpretation of the observed associations. Random sampling is the most essential ingredient of all experimental designs. Besides, there are many devices for increasing the precision of the inference and the calculations.

13.2 Completely Randomized Design (CRD)

This is the simplest type of design. In this design, the homogenous experimental material is divided into different experimental units according to the number of treatment and number of replication. Then, all the treatments are randomly allotted to these experimental units.

Suppose an experiment in which k treatments have been allotted to N units such that i^{th} treatment (i = 1, 2, ….,k) replicated n_i times. Using the information, we wish to test-

H_0: $t_1 = t_2 = …….. = t_i = ….. = t_k$ (No significant difference in treatments)

H_1: $t_1 \neq t_2 \neq …… \neq t_i \neq ….. \neq t_k$ (Significant difference in treatments)

Let y_{ij} (i = 1, 2, ………., k; j = 1, 2, …………., n_i) be the j^{th} observation of i^{th} experimental unit, then the data would appear as follows :

Treatment	Observations							Treatment Total	Treatment Mean
1	Y_{11}	Y_{12}	Y_{13}	...	...	...	y_{1n}	T_1	$\bar{y}_1$
2	Y_{21}	Y_{22}	Y_{23}	...	...	...	y_{2n}	T_2	$\bar{y}_2$
3	Y_{31}	Y_{32}	Y_{33}	...	...	...	y_{3n}	T_3	$\bar{y}_3$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	y_{k1}	y_{k2}	y_{k3}	...	...	...	y_{kn}	T_k	$\bar{y}_k$
Over all total								G	

Here,

$$G = \sum_{i=1}^{k} T_i$$

N = n x k = Total number of observations.

Where, n = Equal number of observations per treatment

Now, for testing H_0, the steps in the analysis of variance procedure for CRD are as follows :

i) Calculate the correction factor (C.F.) as

$$C.F. = \frac{G^2}{N}$$

ii) Find to total sum of squares (TSS)

$$TSS = \sum_i \sum_j yij - C.F. = (y_{11}^2 + y_{12}^2 + ... + y_{1n}^2 + ... + y_{kn}^2) - C.F.$$

iii) Find the sum of squares due to treatment (SST)

$$SST = \sum_{i=1}^{n} \frac{T_j^2}{n} - C.F. = \frac{T_1^2 + T_2^2 + ... + T_k^2}{n} - C.F.$$

iv) Find error sum of squares (ESS) by subtraction. i.e.

ESS = TSS - BSS

v) Prepare the ANOVA (Analysis of variance) table

ANOVA-table for CRD

Sources of variations	Degree of freedom (d.f.)	Sum of Squares (SS)	Mean Sum of Squares (MSS)	Calculated variance ration (F)
Between Treatments	k – 1	SST	$\frac{SST}{k-1} = V_1(\text{say})$	$\frac{V_1}{V_2} = F$
Within treatments or Error	N – k	ESS	$\frac{ESS}{N-k} = V_2(\text{say})$	
Total	N – 1	TSS		

Decision rule

The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for (k-1),(N-k) degree of freedom. After calculating the value of the F-test, the decision about the acceptance or rejection of H_0 is taken in the following manner-

i) If the calculated value of test statistic $F \leq F_{(k-1)(N-k)d.f.}$ (α = 0.05), we accept H_0. Hence, we conclude that the treatments do not differ significantly.

ii) If the calculated value of test statistic $F > F_{(k-1)(N-k)d.f.}$ (α = 0.05), we reject H_0 at 5% level of significance for (k-1),(N-k) degree of freedom. Hence, we conclude that the treatments differ significantly.

iii) If the calculated value of test statistic $F > F_{(k-1)(N-k)d.f.}$ (α = 0.05), we reject H_0 at 1% level of significance for (k-1),(N-k) degree of freedom. Hence, we conclude that the treatments **highly** differ significantly.

Example-1

Twenty one albino rats were randomly selected for an experiment. Three types of feed A, B and C were randomly given to each of seven rats and after a certain period the weight gain (gm) as recorded:

A4 C6 B9 A3 C5 A6 B7 A5 B8 C6 A4 C4

B9 C6 A2 B7 C5 B9 A4 C3 B8

Analyze the data and interpretation the conclusions.

Solution:

Step-1: H_0: These three feeds do not differ significant as regard to body weight.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: Now, arrange the data in tabular form as follows:

Feed	Observations							Total	Average
A	4	3	6	5	4	2	4	28	4.00
B	9	7	8	9	7	9	8	57	8.14
C	6	5	6	4	6	5	3	35	5.00
Total								120	

Step-4: Calculate correction factor as:

$$C.F. = \frac{G^2}{N} = \frac{(120)^2}{21} = 685.7143$$

Step-5: Now, computation of various sums of squares:

Step-6: Calculate Total Sum of Square (TSS)

$$TSS = \sum_i \sum_j yij - C.F. = (y_{11}^2 + y_{12}^2 + ... + y_{km}^2) - C.F. = (4^2 + 3^2 + ... + 5^2 + 3^2) - 685.7143 = 88.2857$$

Step-7: Calculate sum of square due to treatment (SST):

$$SST = \sum_{i=1}^{n} \frac{T_i^2}{n} - C.F. = \frac{T_1^2 + T_2^2 + ... + T_k^2}{n} - C.F. = \frac{28^2 + 57^2 + 35}{7} - 685.7143 = 65.4286$$

Step-8: Calculate error sum of squares (ESS):

ESS = TSS – SST = 88.2857 – 65.4286 = 22.8571

Step-9: Prepare the ANOVA (Analysis of variance) table:

Sources of variations	Degree of freedom (d.f.)	Sum of Squares (SS)	Mean Sum of Squares (MSS)	Calculated variance ration (F)
Between treatment (feed)	k – 1 = 3 - 1=2	65.4286	$\frac{65.4286}{2} = 32.7143$	$\frac{32.7143}{1.2698} = 25.763$
Within treatment (feed) or Error	N – k = 21 – 3=18	22.8571	$\frac{22.8571}{18} = 1.2698$	
Total	N – 1= 21 – 1 = 20	88.4286		

Step-10: Here, degree of freedom

i) for numerator: k – 1 = 3 – 1 = 2

ii) for denominator: N – k = 21 – 3 = 18

Step-11: Table value of F (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$F_{2,18}(0.05)=3.55\, and\, F_{2,18}(0.01)=6.01$$

Step-12: Here, F_{cal} = 25.763 < $F_{2,18}(0.01)$ =6.01, we reject H_0 at 1% level of significance.

Step-13: Hence, these three feeds highly differ significant as regard to body weight.

Step-14: Now, since values of F are found **significant** then **critical difference test (CD-test)** is applied. For calculating the value of critical difference, using following formula :

$$C.D.=t_{18}(0.05)\times\sqrt{\frac{2\times 1.2698}{7}}=2.101\times 0.602=1.27$$

Step-15: The means of different classes of assignable factors are arranged either in ascending order or in descending order of their magnitude.

Feeds	A	B	C
Mean	4.0	8.14	5.0

Step-16: Here, actual difference is found less than the value of C.D. between feed A and feed C, it means that the feed A and feed C do not differ significantly and therefore they are connected with a **straight line**. But, on the other hand, the actual difference is more to the value of C.D. then the two means (i.e. feed B and feed C) are said to be significantly different and they are **not** joined with the straight line.

Conclusion: Hence, feed B is superior then other feeds, where feed A and feed C perform similar.

13.3 Randomized Block Design (RBD)

When experimental materials are heterogeneous, randomized block design is most suitable than CRD. In RBD, the experimental materials (units) are subdivided into homogeneous groups/blocks, each in size equal to the number of treatments. The treatments are then allocated randomly to each of units within each block.

Suppose there are k treatments and n blocks in the rectangular array. Let y_{ij} denotes a cell observation on i^{th} treatment in j^{th} block (i=1, 2, …….,k; j=1, 2, ……., n). The data would appear as follows :

Treatments	Blocks							Total ($T_{i.}$)	Mean ($\bar{y}_i$)
	1	2	3	…	j	…	n		
1	Y_{11}	Y_{12}	Y_{13}	…	y_{ij}	…	y_{1n}	T_1	$\bar{y}_1$
2	Y_{21}	Y_{22}	Y_{23}	…	Y_{2j}	…	y_{2n}	T_2	$\bar{y}_2$
3	Y_{31}	Y_{32}	Y_{33}	…	Y_{3j}	…	y_{3n}	T_3	$\bar{y}_3$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	Y_{i1}	Y_{i2}	Y_{i3}	…	y_{ij}	…	y_{in}	T_i	$\bar{y}_i$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	y_{k1}	y_{k2}	y_{k3}	…	y_{kj}	…	y_{kn}	T_k	$\bar{y}_k$
Total (B_j)	**B_1**	**B_2**	**B_3**	**…**	**B_j**	**…**	**B_n**	**G**	
Mean ($\bar{y}_j$)	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	…	$\bar{y}_j$	…	$\bar{y}_n$		

Here, T_i and $\bar{y}_i$ respectively stand for total and mean of the i^{th} treatment (i = 1, 2, ……, k)

T_j and $\bar{y}_j$ respectively stand for total and mean of the j^{th} block (j = 1, 2, ……, n) and

G stand for total of all the nk observations.

Using the above information, we wish to set up the following two null hypotheses :

a) H_{01}: $t_1 = t_2 = …… = t_i = ….. = t_k$ (No significant difference in treatments)

H_{11}: $t_1 \neq t_2 \neq …… \neq t_i \neq ….. \neq t_k$ (Significant difference in treatments)

b) H_{02}: $b_1 = b_2 = …… = b_j = ….. = b_n$ (No significant difference in blocks)

H_{12}: $b_1 \neq b_2 \neq …… \neq b_j \neq ….. \neq b_n$ (Significant difference in blocks)

Now, for testing H_0, the steps in the analysis of variance procedure for RBD are :

i) Calculate the correction factor (C.F.) as

$$C.F. = \frac{G^2}{nk}$$

ii) Find to total sum of squares (TSS)

$$TSS = \sum_i \sum_j yij - C.F. = (y_{11}^2 + y_{12}^2 + ... + y_{1n}^2 + ... + y_{kn}^2) - C.F.$$

iii) Find the sum of squares due to treatment (SST)

$$SST = \sum_{i=1}^{k} \frac{T_1^2}{n} - C.F. = \frac{T_1^2 + T_1^2 + ... + T_n^2}{n} - C.F$$

iv) Find the sum of squares due to block (SSB)

$$SSB = \sum_{i=1}^{k} \frac{T_1^2}{k} - C.F. = \frac{B_1^2 + B_1^2 + ... + B_n^2}{k} - C.F.$$

v) Find error sum of squares (ESS) by subtraction. i.e.

ESS = TSS - (SST + SSB)

vi) Prepare the ANOVA (Analysis of variance) table

ANOVA -table for RBD

Sources of variations	Degree of freedom (d.f.)	Sum of Squares (SS)	Mean Sum of Squares (MSS)	Calculated (F)	Tabulated (F)
Between treatment	k – 1	SST	$\frac{SST}{k-1} = V_1(\text{say})$	$\frac{V_1}{V_3} = F_1$	$F_{v_1,v_3}(\alpha)$
Between block	n – 1	SSB	$\frac{SSB}{n-1} = V_2(\text{say})$	$\frac{V_2}{V_3} = F_2$	$F_{v_2,v_3}(\alpha)$
Error	(k - 1)(n-1)	ESS	$\frac{ESS}{(k-1)(n-1)} = V_3(\text{say})$		
Total	nk – 1	TSS			

Here, v_1 = (k-1), n_2 = (n-1) and n_3 = (k-1)(n-1)

Decision rule

After forming the above ANOVA table, the decisions regarding H_{01} and H_{02} are taken as :

a) The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for n_1,n_3 degree of freedom. After calculating the value of the F-test, the decision about the acceptance or rejection of H_{01} is taken in the following manner :

   i) If the calculated value of test statistic $F \le F_{V_1 V_3 d.f.}(\alpha = 0.05)$ we accept H_{01}. Hence, we conclude that the treatments do not differ significantly.

   ii) If the calculated value of test statistic $F > F_{V_1, V_3 d.f.}(\alpha = 0.05)$ we reject H_{01} at 5% level of significance for n_1,n_3 degree of freedom. Hence, we conclude that the treatments differ significantly.

iii) If the calculated value of test statistic $F > F_{v_1, v_3 d.f.}(\alpha = 0.01)$, we reject H_{01} at 1% level of significance for n_1,n_3 degree of freedom. Hence, we conclude that the treatments **highly** differ significantly.

b) The test is usually performed at 5% i.e. (α = 0.05) and 1% i.e. (α = 0.01) level of significances for n_2,n_3 degree of freedom. After calculating the value of the F-test, the decision about the acceptance or rejection of H_{02} is taken in the following manner-

i) If the calculated value of test statistic $F \leq F_{v_1, v_3 d.f.}(\alpha = 0.05)$ we accept H_{02}. Hence, we conclude that the blocks do not differ significantly.

ii) If the calculated value of test statistic $F > F_{v_1, v_3 d.f.}(\alpha = 0.05)$, we reject H_{02} at 5% level of significance for n_2,n_3 degree of freedom. Hence, we conclude that the blocks differ significantly.

iii) If the calculated value of test statistic $F > F_{v_1, v_3 d.f.}(\alpha = 0.01)$, we reject H_{02} at 1% level of significance for n_2,n_3 degree of freedom. Hence, we conclude that the blocks **highly** differ significantly.

Example-2:

Test the effect of antioxidant on pesticide induced changes in body weight gain of wistar rats. The data are given as below:

Group	Body weight gain (gm)					Total	Average
	0 day	7 days	14 days	21 days	28 days		
Control	118	122	130	145	152	667	133.4
Treatment-1	120	124	128	140	147	659	131.8
Treatment-2	119	122	124	119	112	596	119.2
Trwatment-3	117	120	128	135	142	642	128.4
Total	474	488	510	539	**553**	2564	
Average	118.5	122.0	127.5	134.75	**138.25**		

Solution:

Step-1: H_0: There is no significant effect of treatments and periods of antioxidant on pesticide induced changes in body weight gain.

Step-2: Select the level of significance, generally, used 5% and 1% level of significance.

Step-3: Calculate correction factor as:

$$C.F.=\frac{G^2}{nk}=\frac{(2564)^2}{5\times 4}=32870.80$$

Step-4: Now, computation of various sums of squares:

Step-5: Calculate Total Sum of Square (TSS)

$$TSS=\sum_i\sum_j yij-C.F.=(y_{11}^2+y_{12}^2+...+y_{kn}^2)-C.F.=118^2+122^2+...135^2+142^2)-328704.80$$

= 2489.20

Step-6: Calculate sum of square due to treatment (SST)

$$SST=\sum_{i=1}^{k}\frac{T_j^2}{n}-C.F.=\frac{T_{11}^2+T_{12}^2+...+T_k^2}{n}C.F.=\frac{667^2+659^2+596^2+642^2}{5}-328704.80=605.20$$

Step-7: Calculate sum of square due to block (SSB)

$$SSB=\sum_{i=1}^{k}\frac{T_j^2}{k}-C.F.=\frac{B_1^2+B_2^2+...+B_n^2}{k}-C.F.=\frac{474^2+488^2+510^2+539^2+553^2}{4}-328704.80=1107.7$$

Step-8: Calculate error sum of squares (ESS):

ESS = TSS – (SST + SSB) = 2489.20 – (605.20 + 1107.70) = 776.30

Step-9: Prepare the ANOVA (Analysis of variance) table:

Sources of variations	Degree of freedom (d.f.)	Sum of Squares (SS)	Mean Sum of Squares (MSS)	Calculated variance ration (F)
Between treatment	k - 1 = 4 - 1=3	605.20	$\frac{605.20}{3}$ = 201.73	$\frac{201.73}{64.69}$ = 3.1184
Between block (period)	n -1 = 5 - 1 = 4	1107.70	$\frac{1107.70}{4}$ = 276.925	$\frac{276.925}{64.69}$ = 4.2808
Error	(n – 1)(k -1) = 12	776.30	$\frac{776.30}{12}$ = 64.69	
Total	nk – 1= 5x4 – 1 = 19	2489.20		

Step-10: Here, degree of freedom (for rows)

i) for numerator: k – 1 = 4 – 1 = 3

ii) for denominator: (n – 1)(k – 1) = (5 – 1)(4 – 1) = 12

Step-11: Here, degree of freedom (for columns)

i) for numerator: n – 1 = 5 – 1 = 4

ii) for denominator: (n – 1)(k – 1) = (5 – 1)(4 – 1) = 12

Step-12: Table value of F – for treatments (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$F_{3,12}(0.05)=3.49\ and\ F_{3,12}(0.01)=5.95$$

Step-13: Here, F_{cal} = 3.1184 < $F_{3,12}(0.05)$ = 3.49, we accept H_0. Hence, there is no significant effect of treatments of antioxidant on pesticide induced changes in body weight gain.

Step-14: Table value of F – for columns (see the table value in "**Statistical table**" prepared by Fisher & Yates)

$$F_{4,12}(0.05) = 3.26 \; and \; F_{4,12}(0.01) = 5.41$$

Step-15: Here, $_{cal}$ = 4.2808 > = 3.26, we reject H_0 at 5% level of significance. Hence, there is significant effect of periods of antioxidant on pesticide induced changes in body weight gain.

Step-16: Now, if the significant difference between the periods, then we apply critical difference (CD) test.

Step-17: For calculating the value of critical difference, using following formula-

$$C.D. = t_{12}(0.05) \times \sqrt{\frac{2 \times EMSS}{r}} = 2.18 \times \sqrt{\frac{2 \times 64.69}{4}} = 12.40$$

Step-18: Now, means of different classes (Periods) in ascending order:

Periods:	0 day	7 days	14 days	21 days	28 days
Means:	118.5	122.0	127.5	134.75	138.25

Step-19: It is concluded that period 14 days, 21 days and 28 days shows the similar effect and superior than other periods, whereas, Period 0 days, 7 days and 14 days shows the similar effect.

❒❒❒

Chapter 14

Computer Basics and Components of Computer

14.1 Computer

Computer is an electronic device designed to accept and store data, process them and produce meaningful results under the direction of detailed step by instructions.

In 1822, **Charles Babbage**, a professor of Cambridge University designed a machine called an "**Analytical Engine**", which could store data and perform arithmetic operations and printing out results. He is known as "**Father of Computer**".

14.1.1 Basic anatomy of computer

The word **Anatomy** of a computer means inner structure of the computer. The study of anatomy of computer is necessary to acquire the knowledge and functioning of its inner components. Thus, essentially a computer consists of the following three main units:

1. Input Unit
2. Processing Unit
3. Output Unit

14.1.1.1 Input Unit

This is the process of entering data and programs into the computer system. The main functions of this unit are to take input from the user and provide the processed output. The instructions and data must be presented to the computer in binary language, the only language understood by the machine, so that it is able to understand and execute these instructions.

14.1.1.2 Central Processing Unit

The Central Processing Unit (CPU) is the brain of the computer. It is also called as microprocessor. It consists of:

i) Arithmetic and logical unit (ALU)

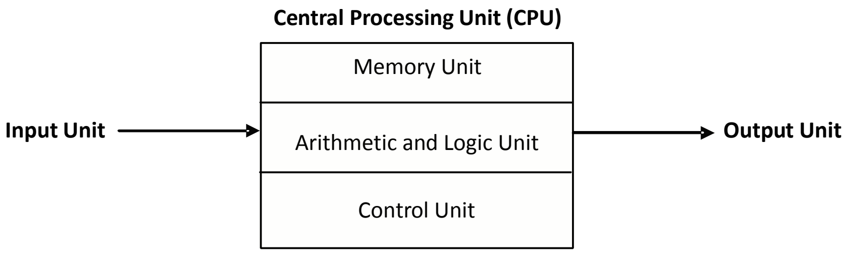
ii) Memory unit (MU)

iii) Control unit (CU)

i) **Arithmetic and logical unit (ALU):** The ALU performs all the arithmetic, comparison and logical operations on the operands. The major operations performed by the ALU are addition, subtraction, multiplication, division, logic and comparison.

ii) **Memory unit (MU):** The memory unit is the medium for storing input and output data, programs and intermediate results. Computer is used to store data and instructions.

iii) **Control unit (CU):** The control unit is the most important part of the CPU as it controls and coordinates the activities of all other units such as ALU, memory unit, input and output units. The process of input, output, processing and storage is performed under the supervision of a unit called **'Control Unit'**. It decides when to start receiving data, when to stop it, where to store data, etc. It takes care of step-by-step processing of all operations inside the computer. Central Processing Unit (CPU), which provides the computer with the processing power to process instructions. The processing power of CPU is measured in terms of Gigahertz (GHz).



Computer is either desktop or vertical tower type. Their chasses (cabinet) are made up of metal or plastic. Now days, a vertical tower type cabinets are used.

14.1.1.3 Output Unit

The **output unit,** as the name implies, provides results to the user. The output unit is a communication link between the computer and the user. This is the process of producing results from the data for getting useful information. Interface with external world for presenting output data. Output may be in the form of hard copy (Paper) or softcopy. Pinter, visual display unit (VDU) terminal, Magnetic disks are the example of output devices.

14.2 Operating System

It is defined as interface between the user and the computer. Also it takes care of memory management, process management, file management and proper control of the system. Features of Operating system are :

i) Hardware management

ii) Command interpretation

iii) Input/output control

iv) Security of the hardware and software

v) Time sharing and process management

vi) Communication

14.2.1 Types of operating systems

We can classify operating system (OS) on the basis of two main factors:

i) On the basis of their text and graphics

ii) On the basis of their user

14.2.1.1 On the Basis of Their Text and Graphics

The operating system that is based on text and the graphics can be divided into the following two types-

a) **Operating system based on Text:** In the text interface the computer user types all his commands on the key board and through this medium he activates them. MS-DOS can be said to be one of the main example of this type of operating system.

b) **Operating system based on graphics:** This interface is also known by the name of **Graphics User Interface (GUI)**. In this process the computer user point out the icons and Menu directions displayed on the display unit screen of the computer which is in the form of various pictures. The GUI symbols are selected with the help of input device "**MOUSE**" and then it sent its commands to the printer for printing jobs.

14.2.1.2 On the Basis of Its User

We can also divide into two parts of operating system based on the interface with the user.

a) **Single User:** The operating system on which only one person can work on the computer at a time. For examples, MS-DOS, Windows, etc.

b) **Multi User:** Multi user operating system is one, which enables more than one person to work at a time. These operating systems are also capable to maintain the network between the computers. For examples, Window-NT, UNIX, etc.

14.3 Windows

14.3.1 What is windows & why is it needed?

Computer understands only machine language. Windows is an operating system that translates user language to machine language. It is an interface between the user and computer. There are two types of interface:

1. **Character User Interface (CUI):** Text or character (like copy, erase etc) is used to establish link between computer and user. Example is DOS.
2. **Graphic User Interface (GUI):** Pictures or graphs are user to establish link between computer and user. One picture is worth thousand words. Example is Windows.

14.3.2 Why is windows a better operating system

- Easy to learn and use.
- Windows and its applications run under protected mode, the crash of single errant program does not automatically crash the operating system of any of other programs.
- Multitasking, Windows is preemptive multitasking operating system means programs running in the background don't significantly degrade the interactive programs that are running in the foreground.
- Windows integrates virtually all your computing tasks and resources– Networks, E-mail, Multimedia, System administration, Printing, fax etc.
- Faster processing (32 bit processing) enhances system efficiency.

14.3.3 Windows V/s DOS

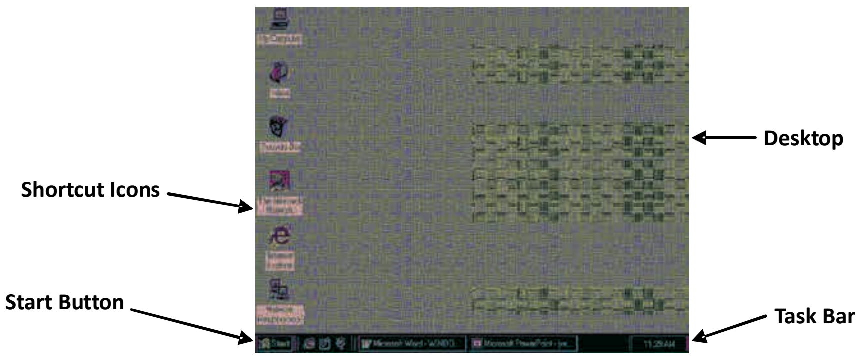
- User friendly – has many menus and dialog boxes to interact with the user.
- Pictorial interface – items represented by pictures, which are easy to remember and comprehend.
- Common menus in all windows, making simple tasks, as opening and closing files very easy.
- Facility to run more than one application or more than one copy of the same application.
- Object linking and embedding.
- Windows accepts long file names, making it possible to give meaningful names to the files or folders.

14.3.3.1 Advantage of GUI

1. User friendly and easy to learn
2. User need remember commands as in DOS.
3. Facilities multitasking and long file names.

14.3.4 Desktop

Computer screen when a window is started is called desktop. On starting windows, you see the Desktop area as shown below:



14.3.5 Shortcut Icons

An icon is a small image representing an object, allows quick access to commonly used programmes, folders or documents.

14.3.6 Task bar

It shows all currently running applications as buttons. Allows shifting from one application to another; it also displays the status and time button. The rest of the desktop is made up of the other items, which are joined together to make one standard bar at the bottom of the screen:

14.3.7 Start button

It allows access to various menu commands. Used for starting a program, find a file or get help, making changes to computer settings. All applications are accessible through the start button.

14.3.8 Status and time

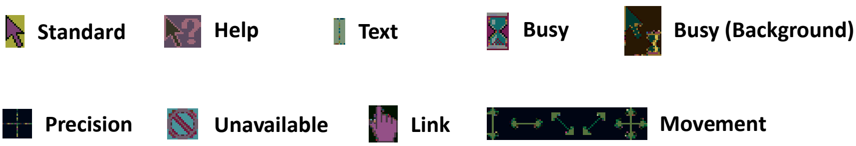
It displays current time for internal clock and provides other status information about the system like print status, etc.

14.3.9 Mouse

At this point The Desktop does not matter and will be explained below. What matters here is that you can identify the desktop.

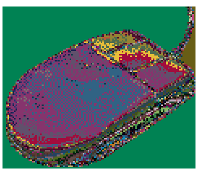
14.3.9.1 Mouse Pointers

Mouse Pointers change according to what your computer is doing and/or what you are doing. They are explained towards the bottom.



14.3.10 Getting started

Before you can get started with the desktop you must first know how to operate the Mouse, which can be tricky at first.



The Mouse

While working in Windows, a mouse is used quite frequently and most of the time left mouse button is used. Most of the windows commands are accessible using the mouse, making windows an easier to use operating system. Before we can proceed, we must know the types of tasks we perform with the mouse:

- **Pointing:** Refers to placing the mouse cursor on the object.
- **Selection:** Pressing the appropriate mouse button so that the object is highlighted.
- **Dragging:** Refers to the holding down of the mouse button and pulling the cursor to another location.

The different types of mouse actions can be classified as:

- **Left Click:** Move the mouse to point at an object press and release left mouse button, this action is used to select an object.
- **Right Click:** Move the mouse to point at an object, press and release the right mouse button. This action is used to display a context relative submenu.
- **Double Click:** Press and release the button twice in quick succession. One can double click to start applications, open and close windows.
- **Drag:** Point at desired object and press the left mouse button, move the mouse without releasing the button, release the mouse button when mouse reaches the final position.

14.4 Shortcut Icons on the Desktop

14.4.1 My computer

It provides the view of computer devices, storage disk and disk drives. It also manages all files and folders.

14.4.2 Network neighborhood

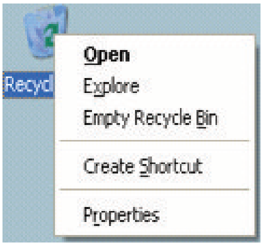
A Network represents a set of computers, which are connected to each other. Double clicking the Network Neighborhood icon allows you to browse through other files and folders of other computers on your network.

14.4.3 Recycle bin

It acts like a wastebasket that helps to keep all our deleted files. Deleted files can be restored from the Recycle bin. Recycle bin has a preset size, which limits the number of files it can store. When the recycle bin is full, it automatically starts removing files that it had stored first. Move the mouse pointer towards to the Full Trashcan icon  until the mouse pointer is hovering over it . Keep the mouse pointer still, whilst over the icon, and then Click (press once) the left mouse button (LMB). This will highlight the icon.

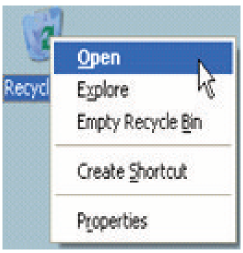
A Click - Pressing the LMB (or RMB) once

If you now **click** (press once) the right mouse button (RMB), whilst the mouse pointer is kept still over the highlighted Full Trashcan icon, you will see a Menu appear.

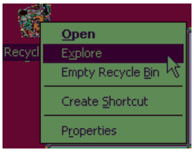


A Click - Pressing the RMB (or LMB) once

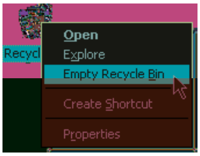
Look at the menu and you will see **Empty Recycle <u>B</u>in** three menu items down. To get to it you highlight the first menu item (**<u>O</u>pen**), by placing the mouse pointer over it, and then you move the mouse pointer down the menu items until you have Empty Recycle <u>B</u>in highlighted. From there, you click the LMB, whilst the mouse pointer is kept still over the highlighted Empty Recycle <u>B</u>in menu item. This then selects Empty Recycle <u>B</u>in, which brings up a Message Requester asking you if you want to delete the item(s) in the Trashcan - Click the LMB on either the **YES** button or **NO** button.



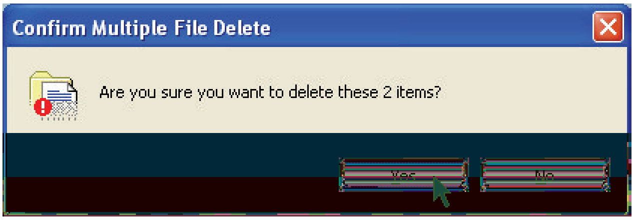
Highlight the Open menu item



Go down the Menu Items



Select (Click with the LMB) the Empty Recycle Bin menu item

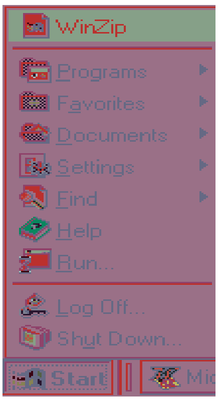


Delete the contents of the Trashcan...YES or NO?

The Trashcan has been emptied

Click anywhere on the desktop to De-Highlight the Trashcan

14.4.4 Components of start menu: On clicking the Start button the following menu appears.



Start menu options

14.4.5 Programs

The Programs menu option gives access to all the applications available in your computer. Each entry represents an application (e.g. Window Explorer) or a group of applications (e.g. Accessories). The group entries have a black arrow at the end and pointing on a group item exposes the items in the group. Clicking on appropriate item can start an application.

14.4.6 Documents

We can quickly open a document we have worked on recently by using this command on start menu. This menu option displays last 15 documents, which have been used.

14.4.7 Settings

Allow you to change setting of the computer, like Background, Printer and Taskbar etc.

14.4.8 Find

Find files or computers quickly through the find command. Allow searching for files by :

1. Specifying filename on wildcard pattern (*, ?).
2. Specifying Date of last modification
3. Specifying type, size of text content.

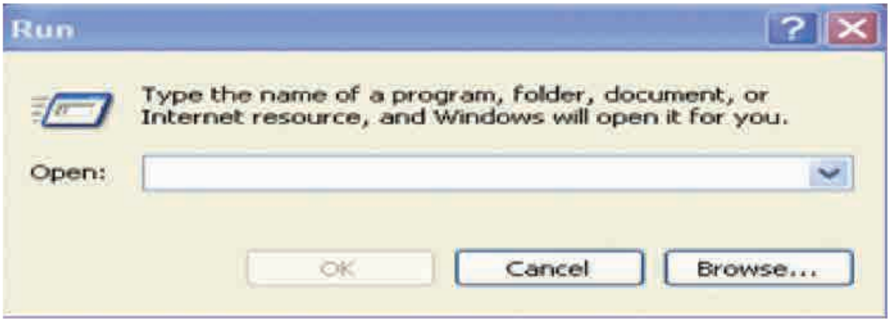
14.4.9 Help

Help available from the start menu provides you with a detailed explanation of all tasks and troubleshooting on windows. The help dialog box has three tabs.

- Contents tab provides help on topics grouped by subject.
- Index tab provides help to find specific topic listed alphabetically.
- Find to type text on which help is sought.

14.4.10 Run

To start a program directly use the RUN command, selecting the Run command displays the Run, in the ‘Open’ List box, type the location and name of the program you want to star. If you don’t remember the location or name of program file, click Browse. To select a program you started recently, click the down arrow in the Open box, click a program in the list, and then click OK.



14.4.11 Shutdown / turn off computer

Shutdown/Turn off computer gives you the following options.

14.4.12 Shutdown / turn off the computer

Use this option, when you want to switch off the computer. Do not turn off your computer until you receive the message "You can now safely turn off your computer".

14.4.13 Restart the computer

Use this option when some program is misbehaving and is halting the system you can use this as a safer alternate than using the computer's reset button.

14.4.14 Restart the computer in Ms-Dos-mode

Gives you the conventional DOS prompt and is used to run DOS based software.

14.4.15 Folders and files

Windows uses the concept of Files and Folders to organize the contents of the computer:

14.4.15.1 Folders

Are the analog for the term directories in conventional DOS. Files and other subfolders are stored in folders. To create the shared folder across a network, click the right mouse button on desired folder and select the sharing option from properties dialog box.

14.4.15.2 Files

Files is basic unit of storage of information or data. Principally two types of files in Windows Program Files (Application Files), User Files, generally referred to as documents.

Icons referring to the application they were made in represent user files. Windows support long file names up to 256 characters. Three-letter extension can be specified to categorize the files. A folder can't contain two files or folders with the same names.

14.5 Windows

Each file or folder is opened in a window. The window of a file or folder consists if the following components.

14.5.1 Title bar

As the name indicates the file bar specifies the title for the window, it represents the name of the file or folder itself, in case it is a file created in a particular application, and then the name of the application is also displayed. The title contains three other buttons viz. three other buttons namely minimize, maximize and close which perform the respective task of minimizing to the taskbar, maximizing to fill the desktop and closing the file, folder or application.

14.5.2 Menu bar

Most windows have a menu with 'File' 'Edit' 'View', and 'Close' as the most common menu options.

14.5.3 Scroll bar

Each window has two bars Vertical and Horizontal, which allow you to view the complete workspace.

14.6 Working with Files and Folders

14.6.1 About files and folders

Double clicking the, **My Computer** icon allows you to see all the storage devices of your computer. Double click the Hard Disk icon, to see content off C: (your Hard Disk). Double clicking any folder allows you to see all files and sub-folders inside. Double clicking a file allows you to open and see the contents. Double clicking is an application to execute it.

14.6.2 View/ quick view of a document

You can view the contents of any file by double clicking on file icon. This will first open the application, in which file was made. You can quickly view a document by right clicking the file icon and selecting quick view. This method uses a special viewer, and does not open the application. Quick view may not be available for files in non-standard software like WordStar.

14.6.3 Creating a new folder

- Click on the folder in which the new folder is to be created, on the left pane of the Explorer window.
- Select file menu, then new folder option.
- Type in new name for the new folder in the area specified.

14.6.4 Moving files/ folders

- Use Explorer to open the folder – containing file to be moved.
- Locate folder into which file is to be moved on the left pane, using scroll bar.
- Drag the file from right pane, on the target folder on the pane.

14.6.5 Copying files /folders

- Use Explorer to open the folder – containing file to be moved.
- Locate folder into which file is to be moved on the left pane, using scroll bar.
- Press the control keys and drag the file from right pane, to target folder on the left pane.
- Release control key only after the entire operation is complete.
- Alternately drag the file with the right mouse. Select "Copy here" in the shortcut menu.

14.6.6 Renaming files / folders

- Click on the folder or file to be renamed.
- Click again, on the name of the file/ folder so the it gets highlighted.
- Edit the name as per your requirement.

14.6.7 Deleting file / folder

- Click on folder or file to be removed.
- Press Delete on keyboard, then click YES to confirm.
- Alternately, clicks Right mouse on the icon and select Delete option.

14.6.8 Restoring / Un-deleting file / folder

- Select Recycle Bin at the bottom of left pane of the explorer window.
- Deleted files appear on the right pane.
- Drag the file back to desired folder.
- Alternately, right click the file and select restore the file back to the original state.

14.6.9 Customizing windows with the control panel

You can choose control panel from settings option of start button or from the control panel icon in the My Computer. Some of the important items are:

14.6.9.1 Regional Settings

Changing the regional settings affects the way program displays dates currencies etc., it has the following Tabs:

Regional Settings	:	Enables selecting language being used.
Number	:	Select default number formats like decimal symbol etc.
Currency	:	Define currency symbol
Time	:	Define time format
Date	:	Specify how date will be displayed.

14.6.9.2 Date & Time Settings

Computer maintains an internal clock. You can change the date of internal clock using this option. Also available by double – clicking the time area of the task bar.

14.6.9.3 Display Settings

Following default options can be changed

- Changing the background

- Changing the screen saver
- Changing the color settings

These options can be changed alternatively, by right clicking on a blank area in the Desktop, and selecting properties.

Some important windows keyboard commands

Command Name	**Purpose/ Remarks**
Ctrl + N	Open a new document
Ctrl + O	Open a existing document
Ctrl + W	Close the current document
Ctrl + S	Save the current document
Ctrl + C	Copies the highlighted matter to store it in clipboard
Ctrl + V	Paste the matter available on the clipboard
Ctrl + X	Cut the highlighted matter to store it in clipboard
Ctrl + B	Bold the highlighted matter
Ctrl + U	Underline the highlighted matter
Ctrl + I	Italic the highlighted matter
Ctrl + A	Select whole document
Ctrl + Shift + =	Superscript the document
Ctrl + =	Subscript the document
Ctrl + P	Prints the current document
Ctrl + Z	Undo an action
Ctrl + Y	Redo an action
Ctrl + F6	Switch between open documents
Ctrl + F	Find What's This ? Button
Ctrl + H	Replace What's This ? Button
Ctrl + G	Go To What ?
Shift + F1	Use the What's This ? Button
F1	Call for help

❒❒❒

Chapter 15

Computer Networks

15.1 Computer Network

It is a group of computers and peripherals which are interconnected to share information and resources. A communication system that supports many users can be called a **Network** or we can say network is a communication system which interconnects many users who have something in common, either with respect to the type of data being sent or to the geographic areas that the users cover. There are three types of computer networks:

15.1.1 Local area network (LAN)

It is a small network (2 to 100 nodes) usually located within a single building or group of buildings belonging to an organization or an institute.

15.1.2 Metropolitan area network (MAN)

A Metropolitan Area Network (MAN) typically provides to distant networks and resources. MAN's are usually high speed fibre-optic networks that connect LAN segments within a specific area, such as a city or district. A MAN consists of cabling and communications equipment that is usually owned and installed by the network owner or company.

15.1.3 Wide area network (WAN)

It provide country-wide or globally connections via telephone lines and satellite link. In a corporate wide network, each department has a local area network (LAN) that allows sharing of files, databases, printers and other peripheral devices. Several departments working together interconnected their networks so that the information may be shared more easily among the departments. It is also called as **International Network** (**Internet)**. It is the worldwide collection of networks and gateways to communicate with one another throughout the world. It links together millions of computer throughout the world. Following are the necessary requirements for the use of Internet:

i) **Personal Computer**

ii) **Modem (short form of modulator and demodulator):** It is used for converting digital signals to analog signals and vice-versa.

iii) **Internet Service Provider (ISP):** The Company that provides you this connection and you are given a password and the user name. These you have to feed into the computer every time you connect to the Internet.

iv) **Telephone Connection:** You must have a telephone connection to connect the Internet.

v) **Browser Software:** To locate and display web pages, you must have Browser software. In other words, it is a software application that displayed the hypertext documents and follow links to other HTML documents on the web. It is also known as web client and is used to locate and display web pages.

15.2 Important Terms used in Internet

15.2.1 Web page

For locating any information on a particular topic, you consult web pages, which contain text, pictures, movies, sound, images and animation or in other words, it is an electronic page on which information is stored on the web and a group of such web pages forms a web site.

15.2.2 World wide web (WWW)

It refers to the Internet's ability to display any kind of written, graphical or audio visual information on the computer from anywhere in the world. For example, information on sports, politics, food, animals, films, etc. can easily be located on different web sites. This information is stored in the digital form as **Web pages**, which you can access with the help of **Web browsers**. Internet Explorer, Netscape Navigator, Opera are some of the famous Web browsers. You can get information on solar system for your requirement through the Internet.

15.2.3 Website

A collection of related web pages is called a web site. It presents information about a particular person, business, organization, etc.

15.2.4 Home page

It is the first page that appears when you visit a particular web site or in other words, starting page of any web site is called **Home Page**.

15.2.5 Uniform resource locator (URL)

It is the address of the web site which you want to access. For example, www.yahoo.com

15.2.6 Chatting

It is like a written phone call. You type the message line and it appears instantly on the computer of other person with whom you want to communicate. He or she replies you back immediately and it also appears on the computer instantly. You can also have voice chatting by which you can converse with your friend or relatives.

15.2.7 Video conferencing

It is an example of real time communication. With the help of webcam and the right software in computer, you can have face to face personal meetings with other persons or group of persons on the internet. Today, some of the critical operations, medical reports or treatments are done by the opinion of expert doctors, etc., is possible through **Video Conferencing**.

15.2.8 E-commerce

You can place an order for the purchase of an item on the Internet if you have a **credit card**. You only have to tell your credit card number and goods will be delivered at your home.

15.2.9 E-mail

E-mail (Electronic mail) is the message communicated through the internet to another computer. This type of message can be sent anywhere in the world and it reaches there in a couple of minutes. It can possible to send graphics, text and even sound and video files to any other internet user on his E-mail address.

❑❑❑

Chapter 16

Entering and Saving Biological Data Through MS-Excel

16.1 Entering Biological Data into Computer

After the forms have been designed, the database in ready for entry of biological data. Biological data is entered one record at a time. To enter the biological data, the user issues a command, which calls up and displays the appropriate form with blank fields. The user then keys in the data for each field in the appropriate spaces. In this manner, the user enters the data for the first record, then for the second record, and so on. In most database systems, the records are automatically assigned a number as they are entered.

16.2 Saving Biological Data into Computer

While entering data into the fields, the tab key or enter key is usually used to move to the next field. Pressing enter or tab key in the last field on the form **saves** the record in the database and moves to a new, blank form for the next record to be entered. In addition to using the tab or enter key to move forward through fields, one can directly go to any field on the form at any time by clicking on it with the mouse.

16.3 Database Concepts Using MS-Excel

MS-Excel is commonly used in data management, but few people can use it truly effectively, for example, many may know how to perform simple sorting or filtering, but most are at a loss to do it to meet specific criteria. Excel is highly capable than most users think. It can make good use of external data, run queries using specific functions, or extract useful data through Access Query. Therefore, we will learn how to get the most from Excel's data management capabilities.

16.3.1 Objectives

For database management, we will be able to:

- Identify the records and fields in a table or list of items
- Use Excel to find specific items in a list
- Use Excel to sort a list over several fields, including a subfield sort
- Use the Excel AutoFilter feature to filter a list using specific criteria

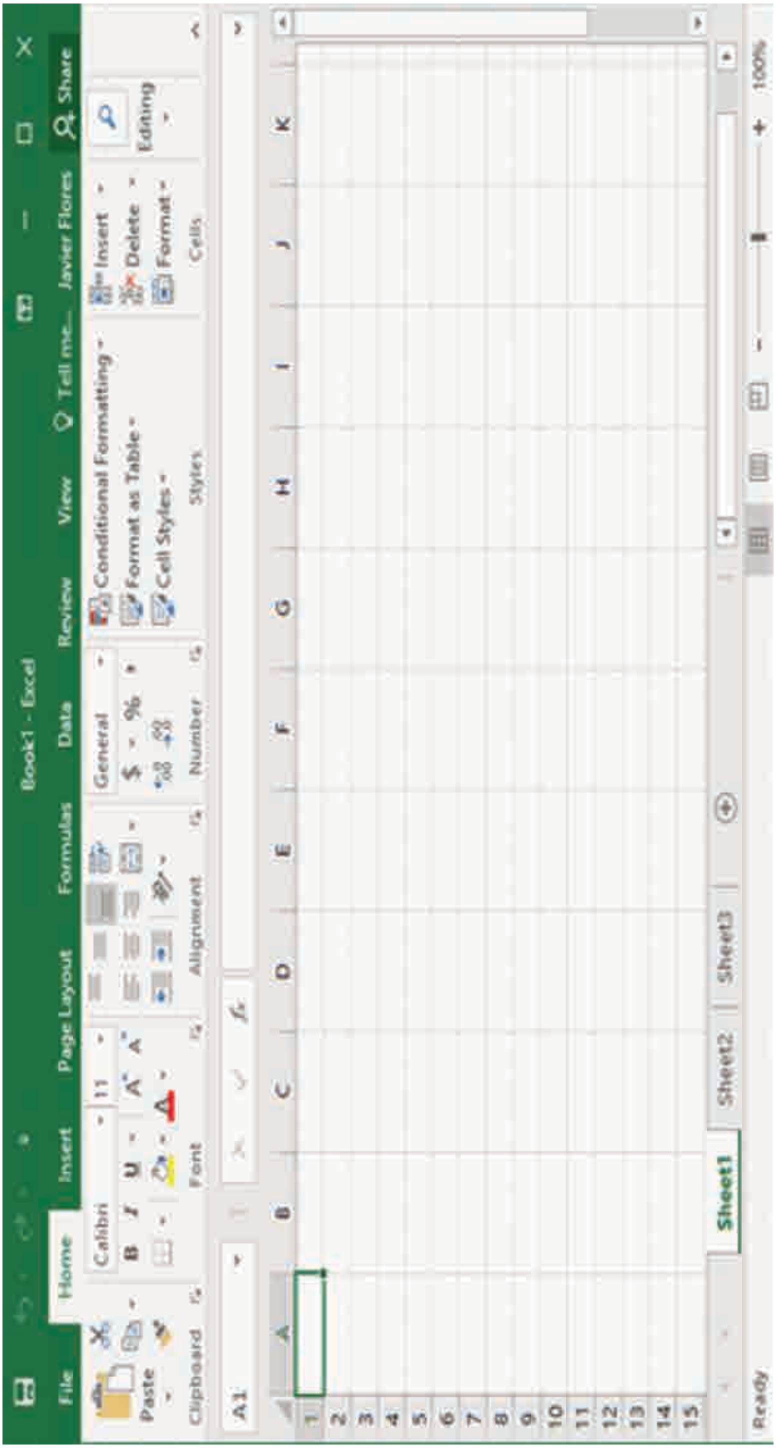
16.4 MS-Excel

Excel is a **spreadsheet program** that allows you to **store**, **organize**, and **analyze information**. While you may believe Excel is only used by certain people to process complicated data, anyone can learn how to take advantage of the program's **powerful features**. Whether you're keeping a budget, organizing a training log, or creating an invoice, Excel makes it easy to work with different types of data.

When you start Excel, a blank workbook appears in the document window. The workbook is the main document used in Excel for storing and manipulating data. A workbook consists of individual worksheet, each of which can obtain data. Initially, each new workbook you create contains **3** worksheets, but you can add more worksheets later.

When you open Excel 2016 for the first time, the **Excel Start Screen** will appear. From here, you'll be able to create a **new workbook**, choose a **template**, and access your **recently edited workbooks**.

From the **Excel Start Screen**, locate and select **Blank workbook** to access the Excel interface. Click the buttons in the interactive below to become familiar with the Excel interface.


Book1 - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Tell me...
Javier Flores
Share
Paste
Clipboard
Calibri
11
Font
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
Editing
A1
A
B
C
D
E
F
G
H
I
J
K
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Sheet1
Sheet2
Sheet3
Ready
100%

The columns are lettered across the top of the document window, beginning with A through Z and continuing with AA through AZ, BA through BZ and so on. The rows are numbered from 1, 2, 3 and so on down the left side of the document window.

The intersections of rows and columns form **cells**, which are the basic units for storing data. Each cell takes its name from this intersection and is referred to as a cell reference. For example, the address of the cell at the intersection of column B and row 5 is referred to as cell **B5**.

At the bottom of each worksheet is a series of sheet tabs, which enable you to identify each worksheet in the workbook. The tabs initially are labeled as Sheet1, Sheet2, and so on.

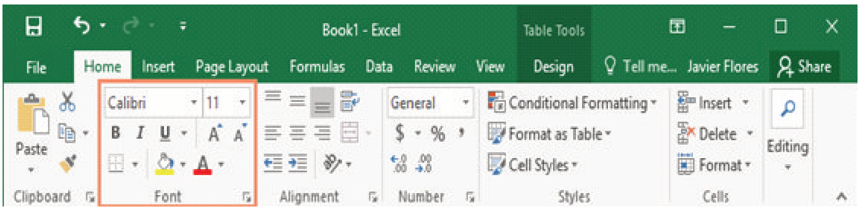
16.5 Working with the Excel Environment

The **Ribbon** and **Quick Access Toolbar** are where you will find the commands to perform common tasks in Excel. The **Backstage view** gives you various options for saving, opening a file, printing, and sharing your document.

16.5.1 The ribbon

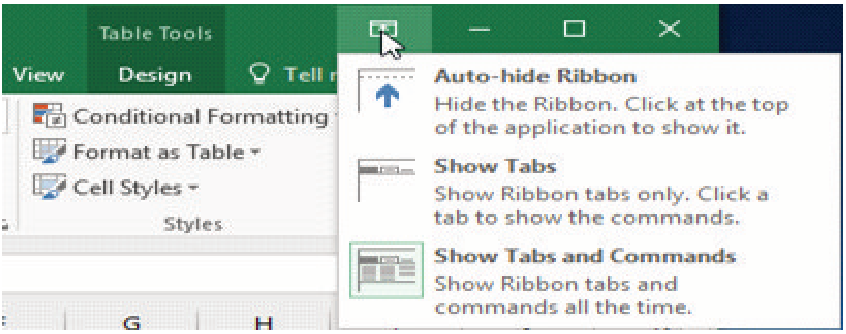
Excel 2016 uses a **tabbed Ribbon system** instead of traditional menus. **The Ribbon** contains **multiple tabs**, each with several **groups of commands**. You will use these tabs to perform the most **common tasks** in Excel.

- Each tab will have one or more groups.



- Some groups will have an arrow you can click for more options.

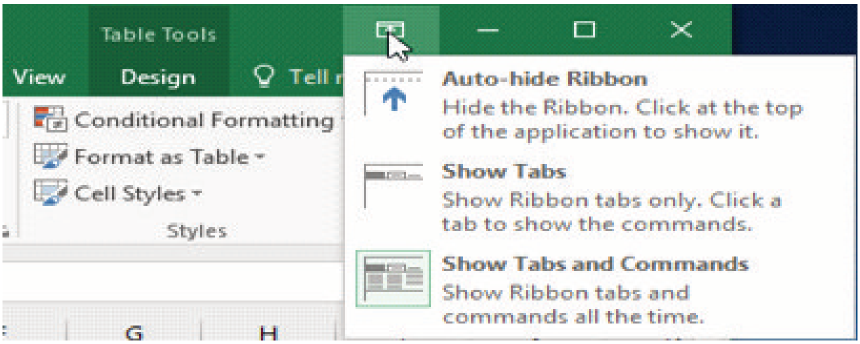
- Click a tab to see more commands.



- Certain programs, such as **Adobe Acrobat Reader**, may install additional tabs to the Ribbon. These tabs are called **add-ins**.

16.5.2 To change the ribbon display options

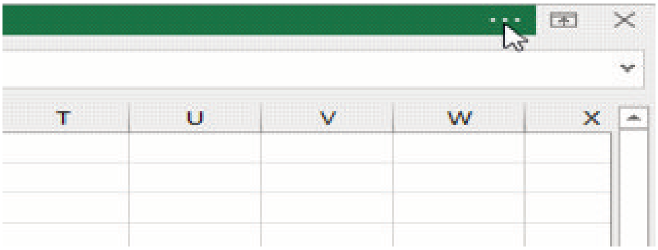
The Ribbon is designed to respond to your current task, but you can choose to **minimize** it if you find that it takes up too much screen space. Click the **Ribbon Display Options** arrow in the upper-right corner of the Ribbon to display the drop-down menu



There are three modes in the Ribbon Display Options menu:

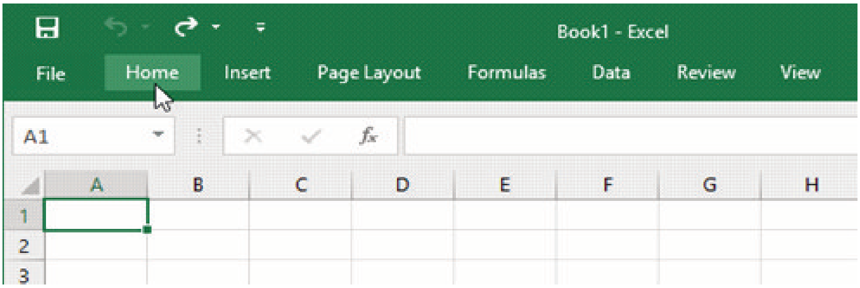
16.5.2.1 Auto-hide Ribbon

Auto-hide displays your workbook in full-screen mode and completely hides the Ribbon. To **show the Ribbon**, click the **Expand Ribbon** command at the top of screen.



16.5.2.2 Show Tabs

This option hides all command groups when they're not in use, but **tabs** will remain visible. To **show the Ribbon**, simply click a tab.



16.5.2.3 Show Tabs and Commands

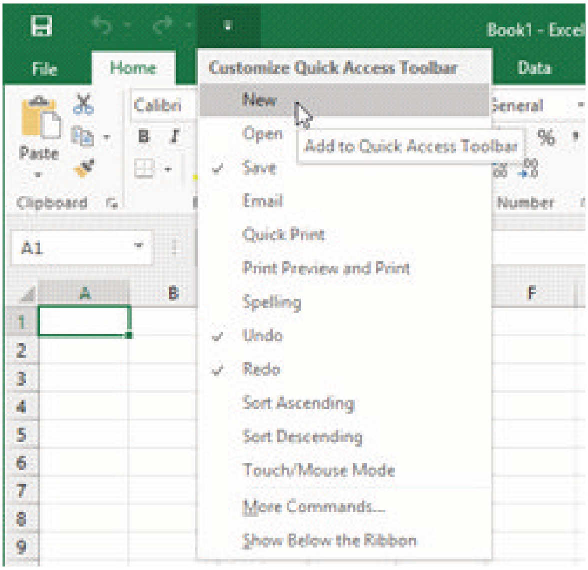
This option maximizes the Ribbon. All of the tabs and commands will be visible. This option is selected by default when you open Excel for the first time.

16.6 The Quick Access Toolbar

Located just above the Ribbon, the **Quick Access Toolbar** lets you access common commands no matter which tab is selected. By default, it includes the **Save**, **Undo**, and **Repeat** commands. You can add other commands depending on your preference.

16.7 To Add Command to the Quick Access Toolbar

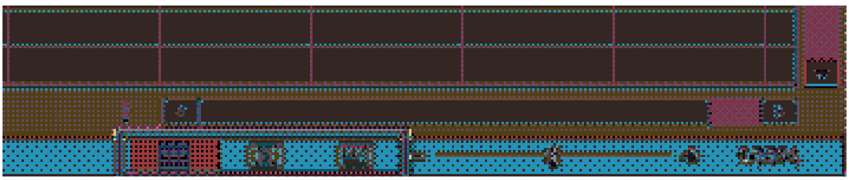
1. Click the **drop-down arrow** to the right of the **Quick Access Toolbar**.
2. Select the **command** you want to add from the drop-down menu. To choose from more commands, select **More Commands**.



3. The command will be **added** to the Quick Access Toolbar.

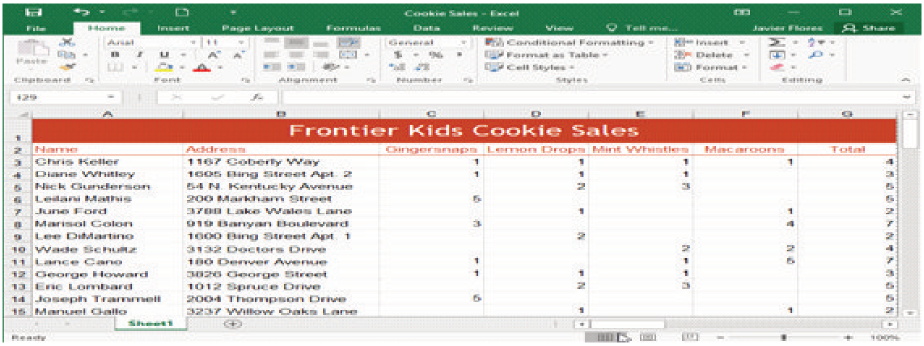
16.8 Worksheet Views

Excel 2016 has a variety of viewing options that change how your workbook is displayed. These views can be useful for various tasks, especially if you're planning to **print** the spreadsheet. To **change worksheet views**, locate the commands in the bottom-right corner of the Excel window and select **Normal view**, **Page Layout view**, or **Page Break view**.



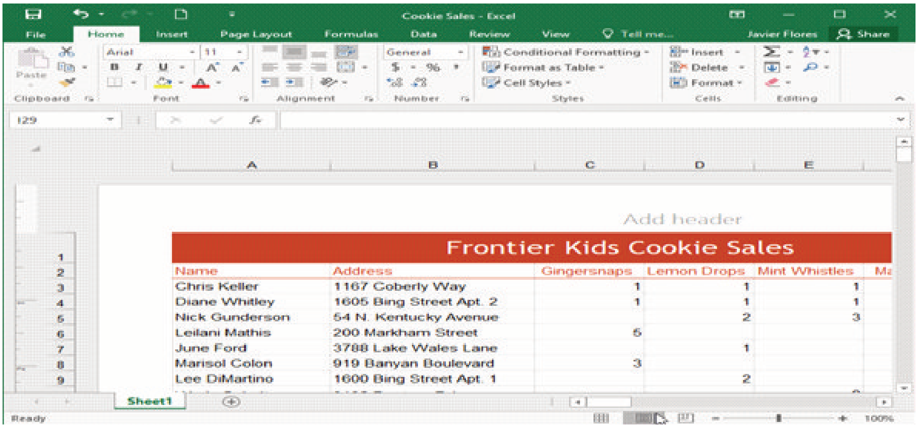
16.8.1 Normal view

Normal view is the default view for all worksheets in Excel.



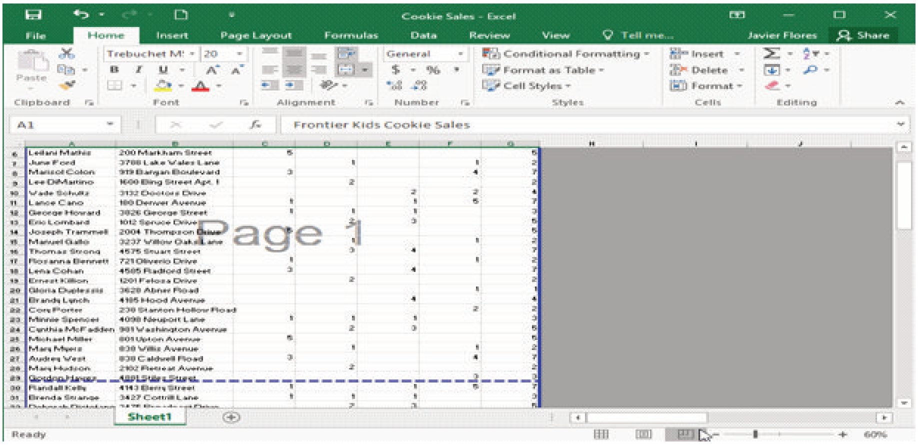
16.8.2 Page layout view

Page Layout view displays how your worksheets will appear when printed. You can also add headers and footers in this view.



16.8.3 Page break view

Page Break view allows you to change the location of page breaks, which is especially helpful when printing a lot of data from Excel.

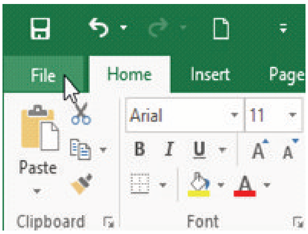


16.9 Backstage View

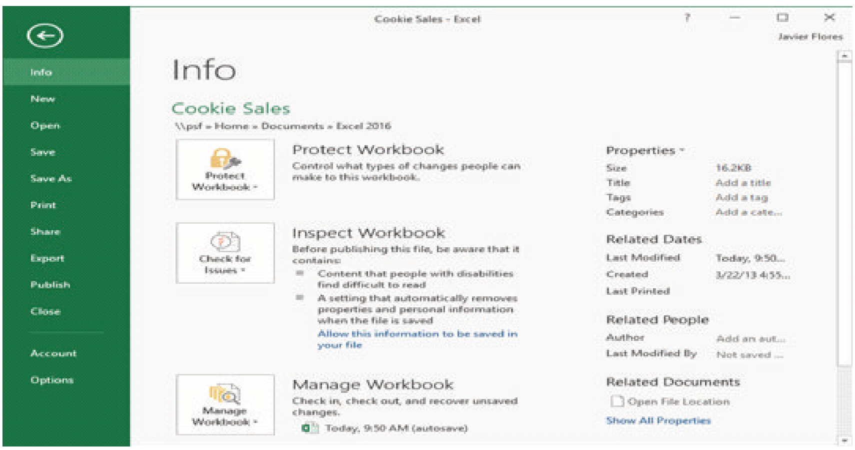
Backstage view gives you various options for saving, opening a file, printing, and sharing your workbooks.

16.10 To Access Backstage View

1. Click the **File** tab on the **Ribbon**. **Backstage view** will appear.

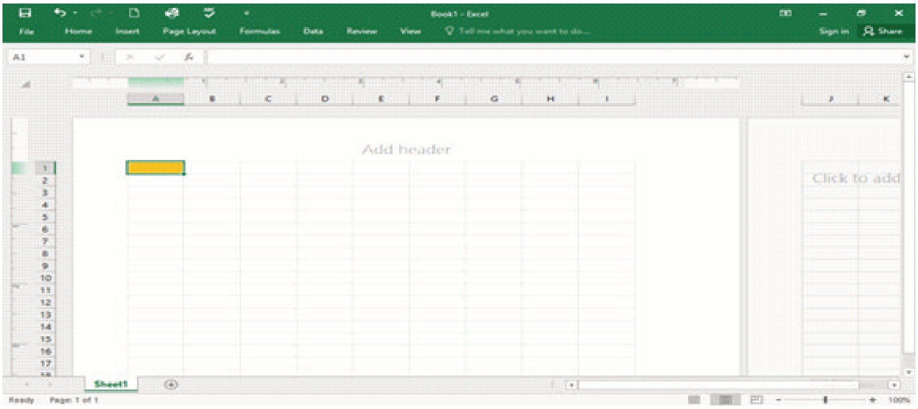


2. Click the buttons in the interactive below to learn more about using Backstage view.



16.11 Starting MS-Excel

1. Open **Excel 2016**.
2. Click **Blank Workbook** to open a new spreadsheet.
3. Change the **Ribbon Display Options** to **Show Tabs**.
4. Using the **Customize Quick Access Toolbar**, click to add **New**, **Quick Print**, and **Spelling**.
5. In the **Tell me bar**, type the word **Color**. Hover over **Fill Color** and choose a **yellow**. This will fill a cell with the color yellow.
6. Change the worksheet view to the **Page Layout** option.
7. When you're finished, your screen should look like this:



8. Change the **Ribbon Display Options** back to **Show Tabs and Commands**.
9. **Close** Excel and **Don't Save** changes.

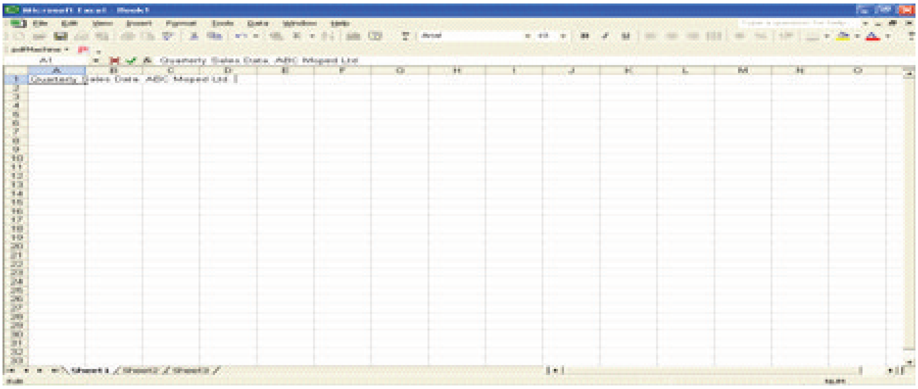
After you activate the cell in which you want to enter data, you can type text, numbers, dates, times, or formulas in the cell. As you type, the data appears in the active cell and in the area above the worksheet called the **Formula Bar**. The active cell displays the insertion point; a blinking bar that indicates where the next character you type will appear. Three small boxes appear between the name box and the insertion point in the formula bar. The first two boxes enable you to reject or accept the data you entered. The reject your entry, click the check box or press **Del or Esc**. To accept your entry, click the check box or press **Enter**. The third box in the formula bar activates the **Edit** formula feature, which simplifies entering formulas in Excel.

16.11.1 Entering text

Text entries consist of alphanumeric characters such as letters, numbers and symbols. You can enter up to 32,000 characters in a single cell, although Excel may not be able to display all the characters if the cell is not wide enough and an entry appears in the cell to its right. When you enter text in a cell, Excel stores entry as text and align it to the left edge of the cell. When you enter data that consists of numbers and text, Excel evaluates the entry to determine its value. If you type an entry such as 2001, Veterinary College makes University, for example, Excel automatically determines that it is a text entry because of the letters.

If you want to enter a number as text, For example, 5656 would be read as a number, but '5656' would be read as a text entry. You can use the inverted comma when you want to enter a number but do not want Excel to interpret it as a value to be used in calculations.

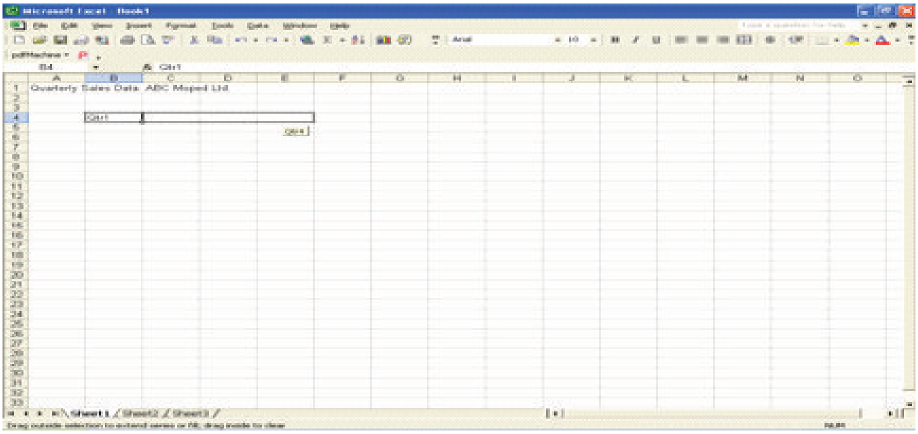
- Go to cell A1
- Enter Quarterly Sales Data: ABC Moped Ltd. in cell A1



16.11.2 Creating a series of text entries

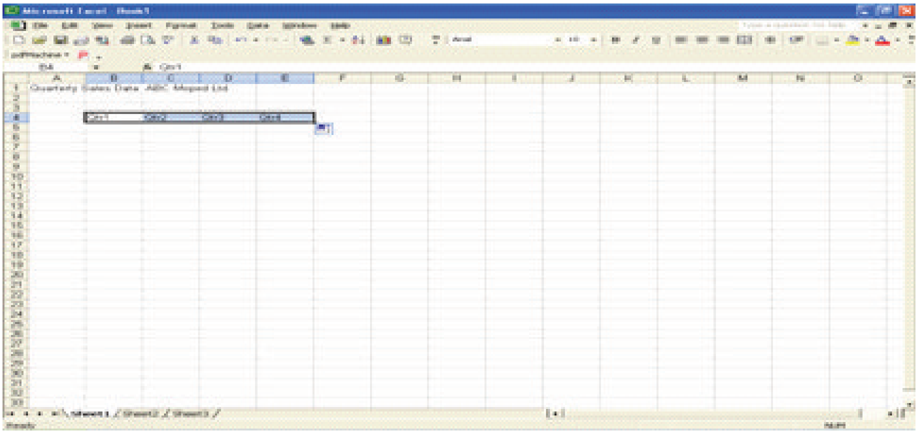
Excel recognizes common text entries, such as days, months and quarterly abbreviations. To fill a range of cells with text entries, follow these steps:

- In cell B4 enter **Qtr1**
- Select the first cell that contains the data i.e. cell B4
- Drag the Auto fill handle over the range of adjacent cells that you want to fill.



- Release the mouse button.

Excel fills the range of selected cells with the appropriate text entries.

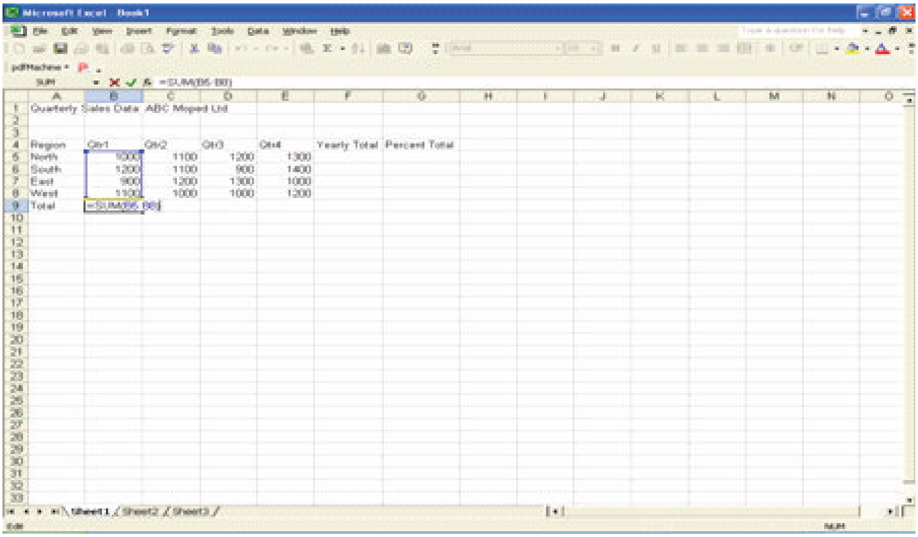


16.11.3 Entering formulas

Most valuable features of Excel are its capabilities to calculate values using formulas. Excel formulas can range from the simple, such as adding a range of values, to the complex, such as calculating the future values of a stream of cash flows. The formula **= SUM (B5:B8)**, for example, adds the values in the range (B5:B8). When the values in these cells change, automatically updates and recalculates the formulas, using the new data in these cells. Excel recognizes a formula in a cell if the entry starts with an equal sign (=) or a plus sign (+) or a minus sign (-). The formula bar continues to show the formula when the cell is the active cell.

16.11.4 Computing totals for the first quarter i.e. Qtr1

- Go to the cell B9 where you want the total to appear
- To enter the formula is complete, press the **Enter** key



16.11.5 Computing totals for the rest of the quarters i.e. Qtr's

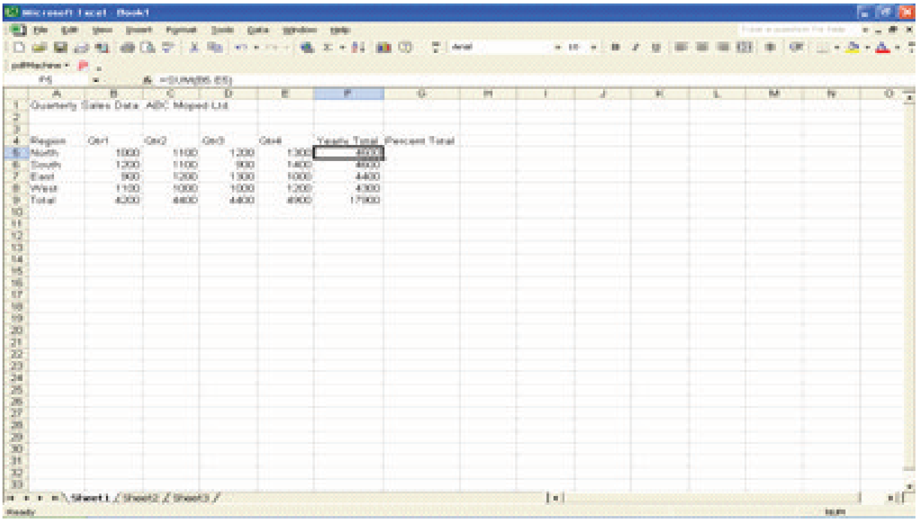
- Copy the formula in cell B9, to the rest of the columns in the same row i.e., the range (C9:E9) through the following steps
- Choose **Copy** from the **Edit** menu
- Select the range from C9 to E9 by highlighting the range
- Choose **Paste** from the **Edit** menu

16.11.6 Computing totals for the North regions

- In the cell F5, enter the formula **= SUM (B5:E5)**

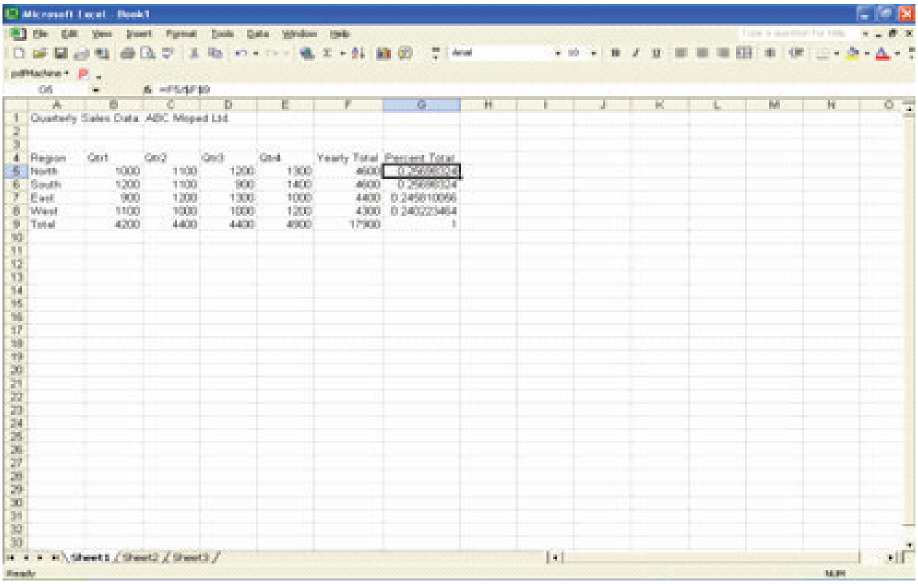
16.11.7 Computing totals for the rest of the regions

- Copy the formula in cell F5, to the rest of the rows in the same column i.e., the range (F6:F9). Your worksheet now contains all the totals.



16.11.8 Computing percent total column

- In the cell G5 enter the formula = F5/F9 (F9 stands for absolute reference)
- Copy the formula from the cell G5 to the range (G6:G9)



16.11.9 Saving your worksheets

- Choose **Save** from the **File** menu
- Type Budget as the name of the file and press the **Enter** key.

Note that the title bar in MS-Excel reflects the file name as **Budget** instead of the default name **Book1**.

16.12 EXCEL'S Range Selection Techniques

As you work with Excel, you will come across many situations in which you will have to select a cell range. Although you can use either mouse or the keyboard to select a range, you will find that the mouse makes the job much easier. The following sections take you through several methods you can use to select a range with a mouse:

16.12.1 Selecting a continuous range with the mouse

A rectangular contiguous grouping of cells is the most common type of range. To use the mouse to select such a range, follow these steps:

- Point the mouse at the upper left hand of the range and then press and hold down the left mouse button.
- With the left mouse button still depressed, drag the mouse pointer to the lower right cell of the range. The cell selector remains around the starting cell, and excel highlights the other cell in the range in the reverse video.
- Release the mouse button. Cells remain selected to show the range.

16.12.2 Selecting a row or a column with the mouse

For a row, click the row's heading, for a column, click the column's heading. If you need to select two adjacent rows or columns, just drag the mouse pointer across the appropriate headings.

16.12.3 Selecting a range in extend mode with the mouse

This method uses the F8 key with the mouse to select a range as follows:

- Click the upper-left cell of the range.
- Press F8. Excel enters the extend mode (you will see EXT in the status bar)

- Click the lower right cell of the range.
- Excel selects the entire range. Press F8 again to turn of the extend mode.

16.12.4 Selecting 3-D ranges

A 3-D range is a range selected on multiple sheets. This is a powerful concept because it allows you to select a range on two or more sheets and then enter data, apply formatting, or give a command, and the operation will affect all the ranges at once.

To select a 3-D range, you first need to group the worksheets you want to work with. To select multiple sheets, you can select any of the following techniques:

- To select adjacent sheets, Click the tab of the first sheet, hold down the Shift key, and click the tab of the last sheet.
- To select noncontiguous sheets, hold down the Ctrl key and click the tab of each sheet you want to include in the group.
- To select all the sheets in a workbook, right click any sheet tab and choose Select All Sheets from the context menu.

When you have selected you sheets, each tab is highlighted and appears in the workbook title bar. To ungroup the sheets, click a tab that is not in the group or select ungroup sheets from the menu appearing on the right-click of the mouse.

With the grouped sheets, you can create your #D range simply by activating one of the grouped sheets and then selecting a range using any of the techniques discussed above. Excel selects the same cells in all the other sheets in the group.

You can also type in a 3D range by hand when entering a formula, for example. The general format is as follows:

16.13 First Sheet: Last Sheet! UL Corner: LR Corner

16.13.1 Using range names

Although ranges let you work efficiently with large group of cells, they have some disadvantages:

a) You can't work with more than one range at a time, each time you want to use a range; you have to redefine its coordinates.

b) A single mistake in defining a range can lead to disastrous results, especially when erasing

c) Range notation isn't intuitive

You can overcome these problems by using range names i.e. you can assign names of up to 255 characters to any single cell or a range of your spreadsheet. Then to include the range in a formula or any other command, you can use the name instead of selecting the range or typing its coordinates. You can create as many range names as you like, and you can even assign multiple names to the same range.

Also a formula like = SUM (August Sales) becomes more intuitive because you don't have to specify the range coordinates.

Named ranges also bring several other advantages to the table like:

1. Names are easier to remember than range coordinates
2. Names don't change when you move a range to another part of the worksheet.
3. Named ranges adjust automatically when you insert or delete columns or rows within the range.
4. Names make it easier to navigate a worksheet. You can use the Go To command to jump to a named range quickly.
5. You can use worksheet labels to create range names quickly.

16.13.2 How to define a range name?

Besides having a length of 255 characters, range names must also follow these guidelines:

a) The name must begin with either a letter or an underscore character. For the rest of the name you can use any combination of characters, numbers or symbols (except space).

b) Don't use cell addresses or any of the other operator symbols because these could cause confusion.

c) To make typing easier, try to keep your names as short as possible while still retaining their meaning.

16.13.3 To define a range

- Select the range you want to name.
- Select Insert/Name/Define. The Define name dialog box appears.

- Enter the Range name. The range selected can also be edited.
- Click the Add button.
- Repeat the step 3 & 4 above for any other ranges you want to name.
- When you are done, click the Close button to return to the worksheet.

16.14 Defining Range Names using Worksheet Text

When you select Insert/Name/Define, Excel sometimes suggests a name for the selected range. Specifically, Excel uses an adjacent text entry to make an educated guess about what you will want to use as a name.

Instead of waiting of Excel to guess, you can tell the program explicitly to use adjacent text as a range name. The following procedure shows you the appropriate steps:

- Select the range of cells, including the text cells that you want to use as the range names.
- Select Insert/Name/Create.
- Excel guesses where the text for the range name is located and activates the appropriate check box. If this is not the one you want, deactivate it and make the proper activation.
- Click **OK.**

16.14.1 Changing a range name

If you need to change the name of one or more ranges, follow one of the two given methods:

- If you have changed some row or column labels, just redefine the range names based on the new text and delete the old names.
- Select Insert/Name/Define. Highlight the name you want to change in the Names in the workbook list, make your changes in the text box and click the Add button.

16.14.2 Deleting a range name

Follow the steps below to remove a range name that is no longer needed:

- Select Insert/Name/Define to display the Define Name dialog box.
- In the names in the workbook list, select the name you want to delete.

- Click the delete button. Repeat if any more ranges are to be deleted.
- Click OK when you are done.

16.15 Working with Charts

One of the best ways to analyze your worksheet data, or get your point across to other people, is to display your data visually in a chart. Excel gives you tremendous flexibility when you are creating charts; it lets you place charts in separate documents or directly on the worksheet itself. Not only that, but you have dozens of different chart formats to choose from.

After you have created a chart and selected the appropriate type of chart, you can enhance the chart's appearance by formatting any of the various chart elements.

16.15.1 Creating a chart

When plotting your worksheet data, you have two basic options. You can create an embedded chart, which sits on top of your worksheet and can be moved, sized and formatted or you can create a separate chart sheet by using the automatic or cut-and paste methods. In both the cases, the charts are linked with the worksheet data. Excel's Chart wizard tool takes you through the steps necessary for setting up a chart as follows:

- Select the cell range you want to plot.
- Either Click the **Chart Wizard** tool on the standard toolbar or select **Insert/Chart**. Excel displays the dialog box.
- Select a chart from the Chart type list and then select a subtype.
- Click the **Next >>** button. Excel displays the Source Data dialog box.
- Use the Data range box to enter the range you want to chart (using mouse or typing the coordinates).

Click the **Next >>** button.

The third dialog box appears.

- This dialog box presents a number of tabs that you can use to format the chart e.g. titles etc. when you're done, click the **Next >>** button to display the final Chart Wizard dialog box.

To insert the chart as a **New Chart Sheet**, activate the **As New Sheet** option and enter a title for the worksheet in the text box provided. If you'd prefer to embed the chart on the existing worksheet, activate the **As Object** in text box and use the drop-down list to choose the sheet you want to choose. When you are done, click **Finish**. Excel inserts the chart.

16.15.2 Converting a series to a different chart type

If you want to create a combination chart not found among the Excel's built-in chart types or if you have a chart formatting you want to preserve, you can easily apply an overlay effect to an existing chart. To do so, follow these steps:

- Activate the chart you want to work with.
- Click the series you want to convert.
- Select **Format/Chart** type to display the Chart Type dialog box.
- In the **Options** group, make sure that the Apply to selection option is activated.
- Select the chart type you want to use for the series and click **OK.** Excel converts the series to the chart type you selected.

16.15.3 Formatting chart axes

Excel provides various options for controlling the appearance of your chart axes. You can hide the axes, set the typeface, size and style of the axis labels, format the axis lines and lick marks, and adjust the axis scale.

16.15.4 Formatting axes patterns

The Patterns tab in the Format Axis dialog box lets you set various options for the axis line and tick marks. Here is a summary:

- **Axis:** These options format the axis line. Select none to remove the line, or select custom to adjust the Style, Color and Weight. The accompanying sample box will show how the line will look.
- **Tick mark type:** These options control the position of the Major and Minor tick marks.
- **Tick mark labels:** These options control the position of the tick mark labels.

16.15.5 Formatting an axis scale

You can format the scale of your chart axes to set things such as the range of numbers on an axis and where the category and value axes intersect. To format the scale, select the Scale tab in the Format Axis dialog box. If you are formatting the value (Y) axis, you can format scale characteristics such as the range of values (Maximum and Minimum), the tick mark units (Major unit and Minor unit), and where the category (X) axes crosses the value axes. Formatting the value axis scale properly can make a big difference in the impact of your charts.

For the category (X) axes, the scale tab options mostly control where the Y axis crosses the X axis and the frequency of categories.

16.15.6 Formatting axis labels

You can change the font, numeric format and alignment of the labels that appear along the axis. To change the label font, select the Font tab in the Format Axis dialog box and then select the font options you want. To change the numeric format of axes labels (assuming that the labels are numbers, dates or times) you have two choices:

- Format the worksheet data series that generated the labels. Excel uses this formatting automatically when its sets up the axis labels.
- Select the Number tab in the Format Axis dialog box and then select a numeric format from the options provided.

To format the alignment of the axis labels, select the Alignment tab in the Format Axis dialog box and then select the option you want from the Orientation group.

16.15.7 Formatting chart data markers

A data marker is a symbol that Excel uses to plot each value (data point). Depending on the type of the marker you are dealing with, you can format the marker's color, patterns, border or style. To begin select the data marker or markers you want to work with.

- If you want to format the entire series, click any data marker in the series, and then select Format/Selected Series. Excel displays the Format Data Series dialog box.
- If you want to format a single data marker, click marker once to select the entire series and then click the marker a second time. Then choose Format/Selected Data Point to display the Format Data Point dialog box

Whichever method you choose, select the Patterns tab to display the formatting options for the series markers. Use the border group to either turn off the border or define the Style, Color and Weight of the marker border. Use the Area section to assign marker colors and patterns.

16.15.8 Displaying and formatting chart gridlines

Adding horizontal or vertical gridlines can make your charts easier to read. For each axis, you can display a major gridline, a minor gridline, or both. The numbers you enter for the axis scales determines the positioning of these gridlines.

16.15.9 Formatting the plot area and background

To format the borders, patterns and colors of both the plot area and the background, follow this procedure:

- Select the plot area or chart background.
- Select either Format/Selected Plot Area or Format/Selected Chart Area to display the appropriate format dialog box.
- In the Patterns tab, select the options you want in the Border and area groups.
- If you're in the Format Chart Area dialog box, you can also select the Font tab to format the chart font.
- Click **OK.**

16.15.10 Adding and formatting a chart legend

If your chart includes multiple data series, you should add a legend to explain the series markers. This makes your chart more readable and easier for others to distinguish each series.

To add a legend, you have two choices:

- Select Chart/Chart Options, activate the legend tab in the Chart Option dialog box and then activate the Show Legend check box.
- Use the Chart toolbar's Legend tool to toggle the legend on or off.

You can format your legends with the same options you used to format the chart text. You can also use the options in the Placement tab to change the position of the legend.

16.16 Basic Rules for Using Excel Formulas

The 5 basic rules to remember as we discuss Excel formulas are:

1. All Excel formulas start with an equal (=) sign. This tells Excel that it is a formula.
2. The answer to the formula displays in the cell into which the formula is entered.
3. Cells are referenced in a formula by their column-row identifier, ie. A1, B2.
4. The symbols for addition, subtraction, multiplication, and division are: + - * /
5. You do not have to enter capital letters in your formula; Excel will automatically capitalize them.

16.16.1 Some basic instructions for using simple math formulas

= A1+A6	this Excel formula adds the contents of cell A1 and A6
= A1+A2+A3	this Excel formula adds the contents of the three cells specified. (See the SUM function for adding multiple numbers)
= A3-A1	this Excel formula subtracts the contents of cell A1 from the contents of cell A3
= B2*B3	this Excel formula multiples the numbers in cells B2 and B3
= G5/A5	this Excel formula divides G5 by A5. (NOTE: If you see the error message #DIV/O! in a cell, you are trying to divide by zero or a null value - which is not allowed.)
= G5^2	this formula tells Excel to square the value in cell G5. The number *after* the caret is the exponent. Likewise, the formula H2^3 would cube the value in cell H2.

We can combine multiple operations in one formula. Make sure you use parentheses where needed or you may not get the correct results (see Order of Operations below). Here are some examples:

= (C1+C3)/C4	This Excel formula adds the value in C1 to the value in C3, and then divides the result by the value in C4

= 4*(A2+A5)+3 This Excel formula adds the contents of A2 and A5, multiples this sum by 4, and then adds 3.

16.17 Mathematical Order of Operations

Remember the Order of Operations by remembering the Phrase **P**lease **E**xcuse **M**y **D**ear **A**unt **S**ally. The letters stand for: Parentheses, Exponents, Multiplication, Division, Addition, and Subtraction. And all operations are carried out from **left** to **right**. Here is how the order is applied:

1. First, any math inside of parentheses is calculated.
2. On the second pass, all exponents are resolved.
3. Then any multiplication OR division is performed.
4. Lastly, any addition OR subtraction is performed.

Note: Even though the Aunt Sally phrase may imply that multiplication is done before division, and addition is done before subtraction, that is not true. They are performed during the same step, or pass, through the formula.

Let's illustrate with a simple formula: **4+2*3**

Since the multiplication must be done first, our expression resolves itself to 4+6=10.

Let's practice with a more complex formula: (2*4)+3^2-8/4

Step 1 - Parentheses: 2*4 = 8. Now our expression reads: 8+3^2-8/4

Step 2 - Exponents: 3^2=9. Now our expression reads: 8+9-8/4

Step 3 - Multiply and Divide: 8/4=2. Now our expression reads: 8+9-2

Step 4 - Add and Subtract: The answer is 15

(an error occurred while processing this directive)

Now test your skill on a complicated formula! 3^(6/3)+(3*3)-2*(6-3)

Step 1 - Parentheses: 6/3=2, 3*3=9, and 6-3=2. So now our formula reads: 3^2+9-2*3

Step 2 - Exponents: 3^2=9. So now our formula reads: 9+9-2*3

Step 3 - Multiply &Divide: 2*3=6. So now our formula reads: 9+9-6

Step 4 - Add and Subtract: 12

16.17.1 Calculating percentages in excel

There are two ways to calculate percentages in Excel, depending on how the worksheet (spreadsheet) is designed.

16.17.2 Display a percent sign in the cell

To calculate a percentage and have the percent sign display in the cell, just **enter the formula in the cell and format the cell as a Percentage.** Example: The formula in cell C2 is**=A2/B2**. If A2=25 and B2=50, then 25÷50=.5 and .5 would normally display. But if we format cell C2 as a Percentage, **50%** will display instead.

As we learned in our beginner's tutorial, **Excel Made Easy**, to format a cell or group of cells, right-click in the cell and click "Format Cells...." Click "Percentage" on the Number tab, indicate the number of decimal points, and click "OK."

A format icon can also be found on the ribbon in newer versions of Excel.

16.17.3 Use a column heading of percent and no percent sign in the cell

Perhaps you want a column to show percentages but don't want the percent sign to display. This is easy. Just multiply each formula in the column by **100** to calculate percentages. **Example:** Column C contains formulas to calculate the percentage of Column A divided by Column B. The formula in cell C2 is **=A2/B2**. Label the column PERCENT, and then change the formula in C2 to **= (A2/B2)*100** and so on down the column. Percentages will display instead of quotients.

Chapter 17
Unsolved Problems

Problems

1. A sample survey regarding number of milch animals in a village of Mathura district was conducted. The number of animals per house were enumerated as given below:

4	5	7	3	9	7	9	2	1	6	3	6
7	8	5	9	3	1	9	2	5	7	3	2
9	3	8	1	8	4	4	6	7	1	9	4
8	6	7	5	7	8	9	5	3	1	8	4
5	6	8	7	2	1	4	6	9	2	3	5
5	6	4	2	8	9	1	2	1	3	4	7
8	9	4	5	7	2	7	5	6	1	2	5

   Construct frequency distribution and find out the number of household having animals less than 4 and more than 6.

2. Following marks were obtained by 60 students in the paper of Biostatistics:

54	76	6	76	9	78	56	45	34	43	54	50
44	47	28	65	3	60	79	85	23	45	43	38
12	2	32	22	47	49	54	50	30	21	27	88
89	94	29	55	31	65	61	37	75	79	42	32
90	98	82	83	86	72	78	66	61	57	52	73

   Construct frequency distribution and find out the number of students having less than 40 marks and more than or equal to 70 marks.

Problems

1. Calculate Arithmetic mean by direct and short-cut method for the haemoglobin (gm/100ml) values in the blood of 6 buffaloes:

Hb (gm/100ml)	:	9.2	10.4	12.7	13.4	11.7	14.2

2. Calculate Arithmetic mean by direct and short-cut method for egg weight (gm) of 5 eggs as given in the following frequency table:

Egg weight (gm)	:	40	44	48	52	56
No. of eggs	:	8	14	22	19	12

3. Calculate Arithmetic mean by direct and short-cut method for the age distribution of cases of foot and mouth (FMD) diseases reported in cattle during a year in Uttar Pradesh.

Age (months) :	60-70	70-80	80-90	90-100
No. of FMD cases :	4	11	26	9

Problems

1. Calculate Geometric mean of the erythrocyte sedimentation rate (ESR) values of blood in 6 dogs as given below:

ESR (mm/hr)	:	7	12	17	14	15	11

2. Calculate Geometric mean for body weight (kg) of 5 goats as given in the following frequency table:

Body weight (kg)	:	40	42	44	46	48
No. of goats	:	7	15	23	18	12

3. Calculate Geometric mean for the age distribution of cases of foot and mouth (FMD) diseases reported in cattle during a year in Uttar Pradesh.

Age (months)	:	60-70	70-80	80-90	90-100
No. of FMD cases	:	4	11	26	9

Problems

1. Calculate Harmonic mean for the duration of pregnancy of 6 cows as recorded below:

Duration of pregnancy (days):	278	292	280	284	288	290

2. Calculate Harmonic mean for body weight (kg) of 5 heifers as given in the following frequency table:

Body weight (kg)	:	180	210	240	270	300
No. of heifers	:	6	15	24	19	11

3. Calculate Harmonic mean for the age distribution of cases of foot and mouth (FMD) diseases reported in cattle during a year in Uttar Pradesh.

Age (months)	:	60-70	70-80	80-90	90-100
No. of FMD cases	:	4	11	26	9

Also compare the relationship between A.M., G.M. and H.M.

Problems

1. Calculate weighted mean for the performance of the students on the basis of pass percentage and the number of students in various professional in B.V.Sc. & A.H. in the Veterinary University, Mathura as follows:

Professional	Pass percentage	Number of students
1st	78	80
2nd	62	58
3rd	83	76
4th	78	54
5th	87	52

Problems

1. Calculate Median of the respiration rate values of 7 horse, recorded as given below:

   Respiration rate (minute) : 9 11 10 12 8 10 8

2. Calculate Median of the pulse rate values of 6 pigs, recorded as given below:

   Pulse rate (minute) : 74 78 72 80 76 75

3. Calculate Median for body weight (gm) of 5 chicks as given in the following frequency table:

Body weight (gm)	:	20	22	24	26	28
No. of chicks	:	4	12	22	17	10

4. Calculate Median for weekly milk yield (litres) of Sahiwal cows as given in the following frequency table:

Weekly milk yield (Litre)	:	0-10	10-20	20-30	30-40	40-50
Number of cows	:	4	11	23	19	13

Problems

1. Calculate Mode of the haemoglobin values of blood in 8 buffaloes as recorded below:

   Haemoglobin (gm/100ml of blood): 8 12 10 12 11 15 14 9

2. Calculate Mode for body weight at first calving of 8 heifers as given in the following frequency table:

Body weight (kg)	:	180	200	220	240	260	280	300	320
No. of heifers	:	5	11	16	26	20	15	11	8

3. Calculate Mode for weekly milk yield (litres) of Sahiwal cows as given in the following frequency table:

Weekly milk yield (Litre)	:	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of cows	:	4	11	19	24	15	12	8

Problems

1. Calculate all the measures of dispersion by direct and short-cut method for the haemoglobin (gm/100ml) values in the blood of 6 buffaloes as recorded below:

Hb (gm/100ml)	:	9.2	10.4	12.7	13.4	11.7	14.2

2. Calculate all the measures of dispersion by direct and short-cut method for egg weight (gm) of 5 eggs as given in the following frequency table:

Egg weight (gm)	:	40	44	48	52	56
No. of eggs	:	6	11	21	18	14

3. Calculate all the measures of dispersion by direct and short-cut method for the age distribution of cases of road accident in cattle during a particular year attended in Veterinary Clinical Complex, Veterinary College, Mathura.

Age (months)	:	50-60	60-70	70-80	80-90	90-100
No. of cases	:	7	16	28	22	12

Problems

Based on One Dimensional Diagram

1. The following data gives the number of milch animals at Livestock Farm, Mathura.

Years	2010	2011	2012	2013	2014	2015	2016
Milch animals	45	50	54	58	64	74	98

   Draw simple bar diagram.

2. Draw multiple bar diagram for the following data:

   Year Cow

Number	**Average**	**Milk yield (litre/day)**
2010-12	45	310
2011-13	50	340
2012-13	54	350
2013-14	58	380
2014-15	64	410

3. Draw sub-divided bar diagram for the following data:

Species	2005-06	2010-11	2015-16
Cattle	48	64	78
Buffalo	42	41	42
Sheep	110	125	180
Goat	200	220	300

Based on Two dimensional diagram

4. The following data gives the cattle population in four different districts of Uttar Pradesh.

Districts	Mathura	Etawah	Jhansi	Balia
Cattle population	45,000	10,000	4,000	15,000

Draw square and circle diagrams.

5. Draw pie diagram for the following data:

Species	2005-06	2010-11	2015-16
Cattle	48	64	78
Buffalo	42	41	42
Sheep	110	125	180
Goat	200	220	300
Poultry	800	1000	1200

Based on Three dimensional diagram

6. Draw cube diagram for the following data gives the buffalo population in four different countries.

Country	India	Pakistan	Nepal	Bangladesh
Buffalo population (thousands)	1000	216	8	64

Problems

1. Construct historigram for the following data is related to the expenditure and income statement in different years at a dairy farm.

Years	2001	2002	2003	2004	2005	2006	2007	2008
Expenditure (Rs. in lakh)	21	25	28	32	35	38	48	54
Income (Rs. in lakh)	5	6	8	11	14	15	21	25

2. Construct histogram, frequency polygon, frequency curve for the following frequency distribution:

Dry period (days) :	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	80-90	90-95
No. of crossbred cows:	5	9	13	17	24	21	18	12	7	4

Also find the value of mode by graphically.

3. Construct cumulative frequency curve for the following frequency distribution:

Milk yield (litre) :	0-2	2-4	4-6	6-8	8-10	10-12	12-14	12-16	16-18	18-20
No. of crossbred cows:	4	7	13	17	24	20	18	14	8	5

Also find the value of median by graphically.

Problems

1. From the data on age (years) and blood pressure (mm of Hg) of six adult males are given as follows:

Age (years) :	40	55	73	44	72	76
Blood pressure (mm of Hg):	124	128	165	135	130	148

i) Draw scattered diagram and give interpretation of the results.

ii) Calculate correlation coefficient between age and blood pressure.

iii) Calculate regression coefficient of age on blood pressure.

iv) Calculate regression coefficient of blood pressure on age.

v) Estimate the value of age based on the blood pressure 140 mm.

vi) Estimate the value of blood pressure based at the age of 60 years.

vii) Prove that coefficient of correlation is the geometric mean of two regression coefficients.

Problems

1. A surgeon transplants the kidney in 200 cases and succeeds in 170 cases. Calculate the probability of success after operation.
2. The probability that a man aged 60 years will survive 10 years is 2/5 and a women aged 50 years surviving 10 years is 3/4. What are the chances that they both will survive 10 years?
3. The probability that a man suffers from arthritis is 4/9 and the probability that he suffers from hypertension is 5/7. What is the probability of getting at least one of the disease?

Problems

1. If X is a normal variable with the mean m = 25 and standard deviation s = 5, find the values of Z_1 and Z_2 such that -

   $P(20 < X < 30) = P(Z_1 < Z < Z_2)$

   Here, Z is a standard normal variable.
2. If Z is a standard normal variable. Find the following probabilities or areas

   i) $P(1.2 < Z < \infty)$ ii) $P(0 < Z < 1)$ iii) $P(-1 < Z < 0)$

   iv) $P(-1 < Z < 1)$ v) $P(-1 < Z < 1.25)$ vi) $P(1 < Z < 2.5)$

   viii) $P(-2 < Z < -1)$
3. Of a large group of goat, 9% are less than 45 kg in weight and 90% are less than 70 kg. Assuming a normal distribution, find the mean weight and standard deviation.

Problems

1. A sample of 400 individuals is found to have mean height of 160.3 cms. Can it be reasonably regarded as a sample from a large population with mean height 160 cms with a standard deviation 3.0 cms?
2. In an experiment, two diets are compared on 40 and 50 calves. The average increase in weights due to two diets A and B are 6kg and 5kg with standard deviations 1.0kg and 1.3kg, respectively. Test whether there are significant differences in their mean weights.
3. In a survey of 240 people 105 were found to be regular smokers. Can we conclude from sample data that the proportion of smokers in the sample population is different from 50 percent?

4. In a random sample of 1000 persons from city-A, 400 are found to be consumers of milk. In another sample of 800 persons from city-B, 350 are found to be consumers of milk. Do these data reveal a significant difference between city-A and city-B, so far as the proportion of milk consumers is concerned?

Problems

1. Six buffaloes were randomly selected from a herd and the lactation yield (litre) was recorded as given below:

   Lactation yield (litre): 950 1100 1250 980 1280 1050

   Test whether the average lactation yields are significantly different from the population mean of 900 litres?

2. An experimental ration was given to 8 cows, after a certain period. The change in daily milk yield (litres) in comparison to normal ration as given earlier, was recorded as below:

   Daily milk yield (litres): 3.5 -1 1.5 0 0.5 -2 2.5 3

   Test whether the experimental ration had a significant effect in changing the milk yield?

3. Two new types of rations were fed to pigs. Five pigs were fed Type-A ration and another 7 pigs were fed Type-B ration. The gain in weight (kg) was recorded as given below:

Type-A :	15	22	25	16	28		
Type-B :	18	21	20	15	22	25	16

   Test whether the effect of two rations differed significantly?

4. The milk yield (litre) of six cows in first and second lactation are given below:

First Lactation (litre) :	950	1100	1000	1200	900	1150
Second Lactation (litre) :	970	1050	1075	1250	940	1200

   Test whether there was significant difference in their milk yield of second lactation over first lactation?

5. From the data on milk yield (litre) and butter fat (%) of six buffaloes are given as follows:

Milk yield (litre):	8	12	7	5	2	15
Butter fat (%) :	5.7	4.8	6.0	5.5	7.0	4.5

Test the significance of:

i) coefficient of correlation between milk yield and butter fat?

ii) regression coefficient of milk yield on butter fat?

iii) regression coefficient of butter fat on milk yield?

Problems

1. Lactation period (days) of 5 Bhadawari buffaloes are given below:

   Lactation period (days): 280 240 270 190 210

   Test whether these buffaloes belong to the population having variance 400 days?

2. The number of Hariana cows calved during different months in a particular year at a Dairy Farm is given below:

Month:	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
No. of calving:	50	57	62	40	26	24	18	14	16	22	34	35

   Test whether the calving is normally distributed round the year?

3. Grading of 300 semen samples by two techniques 'A' and 'B' was done. The results are given below:

Techniques	**Below Average**	**Average**	**Above Average**
A	66	34	67
B	60	35	38

   Would you say that the grading techniques are significantly different?

4. Can vaccination be regarded as preventive measure of small pox as evidenced by the following data:

Group	**Affected**	**Non-affected**
Vaccinated	12	13
Non-vaccinated	18	17

5. In an experiment on immunization of dogs against tuberculosis, the following results were obtained:

	Affected	**Non-affected**
Inoculated	12	26
Not inoculated	16	4

   Examine the effect of vaccine in controlling the incidence of disease?

6. The theory predicts the proportion of beans, in the four groups A, B, C and D should be 9 : 3 : 3 : 1. In an experiment among 1600 beans, the numbers in the four groups were 882, 313, 287 and 118. Does the experimental result support the theory?

Problems

1. Compare the performance of four different types of cross bred cows on the basis of their daily milk yield (litre) as given below:

Breed

A	11	8	14	12	15
B	17	15	18	19	20
C	22	24	28	26	30
D	14	19	21	24	25

2. Five breeds of cattle A, B, C, D and E were given on three different rations. Gains in weights (lb) over a given period were recorded.

Breeds

		A	B	C	D	E
	1	47	54	41	46	52
Ration	2	48	42	34	36	32
	3	50	48	38	36	44

Test whether there is significant difference between breeds and between rations?

Problems

1. Fifteen day old chicks were randomly selected for an experiment from same hatch. Three types of feed A, B and C were randomly given to each of five chicks and after a certain period the weight gain (gm) as recorded:

A4 C8 B9 A3 C6 A6 B7 A5 B5 C7 A4 C8 B4 C9 B8

Analyze the data and interpretation the result.

2. A livestock owner grouped into four his 24 cows of 6 breeds and fed them with 4 rations viz. A, B, C and D for a fortnight. The milk yield (litre) are given as below:

Ration	Breed					
	I	II	III	IV	V	VI
A	18	22	20	26	22	21
B	24	20	26	30	27	25
C	25	22	24	21	27	24
D	27	26	21	20	18	22

Test whether there is significant difference between rations as well as breeds?

Problems

1. What is meant by a computer system? Draw a block diagram to illustrate the basic organization of a computer system and explain the functions of the various units.
2. What is an input interface? How does it differ from an output interface?
3. Give the process of starting Windows.
4. Give the different steps to exit Windows.

Problems

1. What is meant by Computer Network? Describe their uses in daily life.
2. What are the advantages of computer network?

Problems

Session-I

1. Create and entering data in three different columns in Excel sheet in practice.
2. Save worksheet in a workbook called student.xls
3. Enter data in 10 rows.
4. Create a duplicate copy of the active sheet after sheet.
5. Find out the student who has the highest mark.
6. Rename the sheet a "Student".
7. Exit the Excel file.

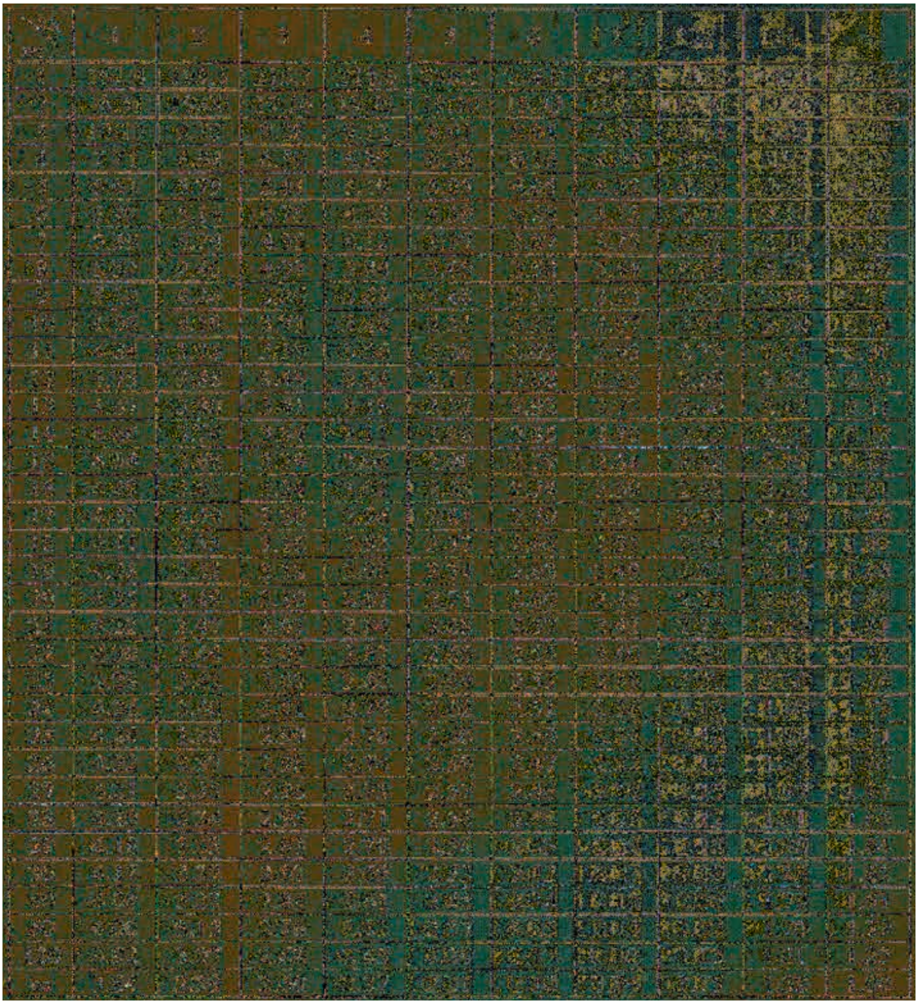
Session-II

1. Opening the existing Excel file.
2. Rename the excel sheet by "Marks Sheet".
3. Calculate the marks (%) at the end of marks entered in first column by using formula
4. Calculate the marks (%) at the end of marks entered in second and third columns by using the above formula
5. Calculate Grade Point (GP) of each columns by using the formula Calculate Overall grade point (OGPA) by using the formula.

❐❐❐

Appendix

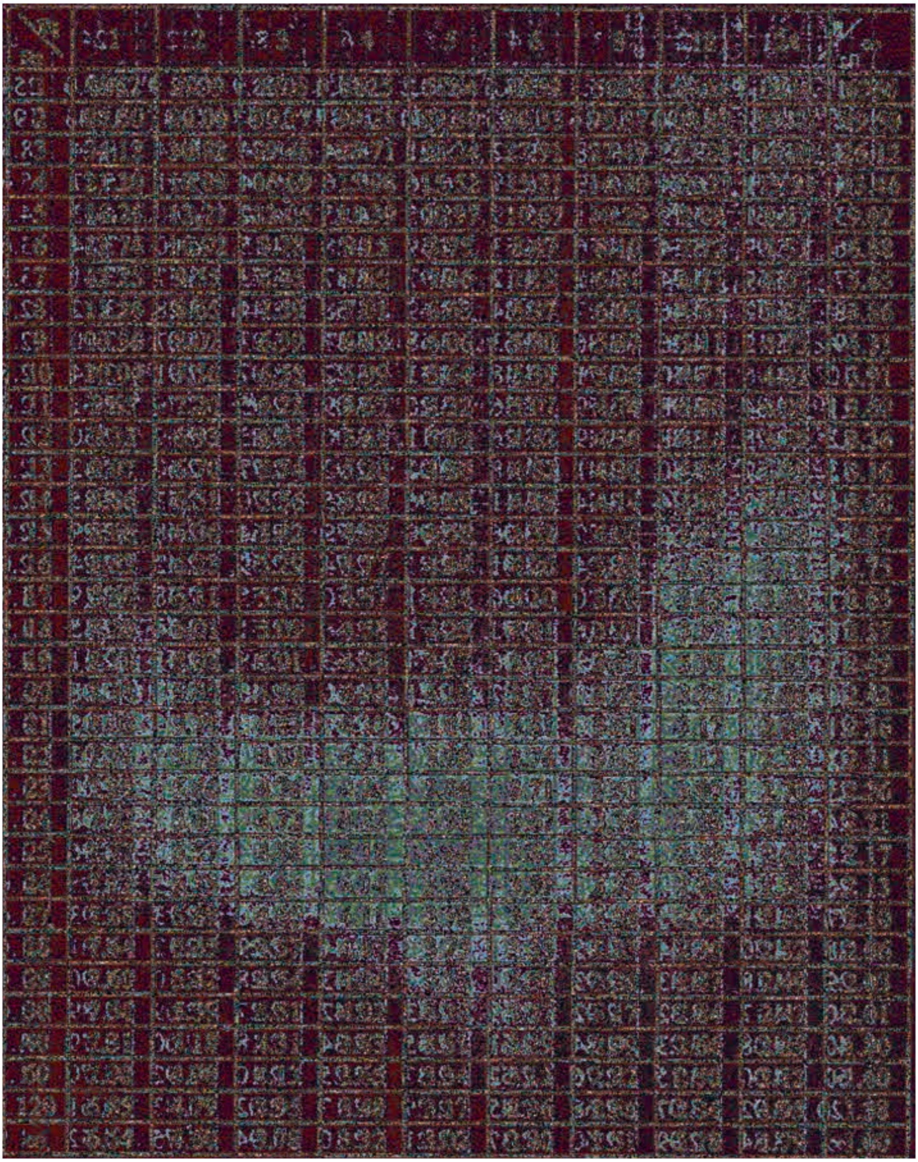
Table 3. (F-Varance ratio) at 5% leverl of significance



[illegible]

Source: Fisher and Yates Statistical Tables

Table 4. (F-Varance ratio) at 1% leverl of significance



Source: Fisher and Yates Statistical Tables

Table 7. Values of χ^2 and t at 1%, 5% and 10% leverl of significance

d.f.	Significance level of t			Significance level of χ^2		
	.01	.05	.10	.01	.05	.10
1	63.66	12.71	6.31	6.63	3.84	2.71
2	9.92	4.30	2.92	9.21	5.99	4.61
3	5.84	3.18	2.35	11.34	7.81	6.25
4	4.60	2.78	2.13	13.28	9.49	7.78
5	4.03	2.57	2.02	15.09	11.07	9.24
6	3.71	2.45	1.94	16.81	12.59	10.64
7	3.50	2.36	1.90	18.50	14.07	12.02
8	3.36	2.31	1.86	20.09	15.51	13.36
9	3.25	2.26	1.83	21.67	16.92	14.68
10	3.17	2.23	1.81	23.21	18.31	16.00
11	3.11	2.20	1.80	24.72	19.68	17.28
12	3.06	2.18	1.78	26.22	21.03	18.55
13	3.01	2.16	1.77	27.69	22.36	19.81
14	2.98	2.14	1.76	29.14	23.68	21.06
15	2.95	2.13	1.75	30.58	25.00	22.31
16	2.92	2.12	1.75	32.00	26.30	23.54
17	2.90	2.11	1.74	33.41	27.59	24.77
18	2.88	2.10	1.73	34.80	28.87	25.99
19	2.86	2.09	1.73	36.19	30.14	27.20
20	2.84	2.09	1.72	37.57	31.41	28.41
21	2.83	2.08	1.72	38.93	32.67	29.62
22	2.82	2.07	1.72	40.29	33.92	30.81
23	2.81	2.07	1.71	41.64	35.17	32.01
24	2.80	2.06	1.71	42.98	36.42	33.20
25	2.79	2.06	1.71	44.31	37.65	34.38
26	2.78	2.06	1.71	45.64	38.88	35.56
27	2.77	2.05	1.70	46.96	40.11	36.74
28	2.76	2.05	1.70	48.28	41.34	37.92
29	2.76	2.04	1.70	49.59	42.56	39.09
30	2.75	2.04	1.70	50.89	43.77	40.26
70				100.42	90.53	85.53
∞	2.58	1.96	1.64			

Source: Fisher and Yates Statistical Tables

Table III. The Normal Probability Integral

x		0	1	2	3	4	5	6	7	8	9
0·0	0·	50000	49601	49202	48803	48405	48006	47608	47210	46812	46414
0·1		46017	45620	45224	44828	44433	44038	43644	43251	42858	42465
0·2		42074	41683	41294	40905	40517	40129	39743	39358	38974	38591
0·3		38209	37828	37448	37070	36693	36317	35942	35569	35197	34827
0·4		34458	34090	33724	33360	32997	32636	32276	31918	31561	31207
0·5		30854	30503	30153	29806	29460	29116	28774	28434	28096	27760
0·6		27425	27093	26763	26435	26109	25785	25463	25143	24825	24510
0·7		24196	23885	23576	23270	22965	22663	22363	22065	21770	21476
0·8		21186	20897	20611	20327	20045	19766	19489	19215	18943	18673
0·9		18406	18141	17879	17619	17361	17106	16853	16602	16354	16109
1·0		15866	15625	15386	15151	14917	14686	14457	14231	14007	13786
1·1		13567	13350	13136	12924	12714	12507	12302	12100	11900	11702
1·2		11507	11314	11123	10935	10749	10565	10383	10204	10027	98525
1·3	0·0	96800	95098	93418	91759	90123	88508	86915	85343	83793	82264
1·4		80757	79270	77804	76359	74934	73529	72145	70781	69437	68112
1·5		66807	65522	64255	63008	61780	60571	59380	58208	57053	55917
1·6		54799	53699	52616	51551	50503	49471	48457	47460	46479	45514
1·7		44565	43633	42716	41815	40930	40059	39204	38364	37538	36727
1·8		35930	35148	34380	33625	32884	32157	31443	30742	30054	29379
1·9		28717	28067	27429	26803	26190	25588	24998	24419	23852	23295
2·0		22750	22216	21692	21178	20675	20182	19699	19226	18763	18309
2·1		17864	17429	17003	16586	16177	15778	15386	15003	14629	14262
2·2		13903	13553	13209	12874	12545	12224	11911	11604	11304	11011
2·3		10724	10444	10170	99031	96419	93867	91375	88940	86563	84242
2·4	$0{\cdot}0^{2}$	81975	79763	77603	75494	73436	71428	69469	67557	65691	63872
2·5		62097	60366	58677	57031	55426	53861	52336	50849	49400	47988
2·6		46612	45271	43965	42692	41453	40246	39070	37926	36811	35726
2·7		34670	33642	32641	31667	30720	29798	28901	28028	27179	26354
2·8		25551	24771	24012	23274	22557	21860	21182	20524	19884	19262
2·9		18658	18071	17502	16948	16411	15889	15382	14890	14412	13949
3·0		13499	13062	12639	12228	11829	11442	11067	10703	10350	10008
3·1	$0{\cdot}0^{3}$	96760	93544	90426	87403	84474	81635	78885	76219	73638	71136
3·2		68714	66367	64095	61895	59765	57703	55706	53774	51904	50094
3·3		48342	46648	45009	43423	41889	40406	38971	37584	36243	34946
3·4		33693	32481	31311	30179	29086	28029	27009	26023	25071	24151
3·5		23263	22405	21577	20778	20006	19262	18543	17849	17180	16534
3·6		15911	15310	14730	14171	13632	13112	12611	12128	11662	11213
3·7		10780	10363	99611	95740	92010	88417	84957	81624	78414	75324
3·8	$0{\cdot}0^{4}$	72348	69483	66726	64072	61517	59059	56694	54418	52228	50122
3·9		48096	46148	44274	42473	40741	39076	37475	35936	34458	33037
4·0		31671	30359	29099	27888	26726	25609	24536	23507	22518	21569
4·1		20658	19783	18944	18138	17365	16624	15912	15230	14575	13948
4·2		13346	12769	12215	11685	11176	10689	10221	97736	93447	89337
4·3	$0{\cdot}0^{5}$	85399	81627	78015	74555	71241	68069	65031	62123	59340	56675
4·4		54125	51685	49350	47117	44979	42935	40980	39110	37322	35612
4·5		33977	32414	30920	29492	28127	26823	25577	24386	23249	22162
4·6		21125	20133	19187	18283	17420	16597	15810	15060	14344	13660
4·7		13008	12386	11792	11226	10686	10171	96796	92113	87648	83391
4·8	$0{\cdot}0^{6}$	79333	75465	71779	68267	64920	61731	58693	55799	53043	50418
4·9		47918	45538	43272	41115	39061	37107	35247	33476	31792	30190

LOGARITHMS

	0	1	2	3	4	5	6	7	8	9	Mean Differences 1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170						5	9	13	17	21	26	30	34	38
						0212	0253	0294	0334	0374	4	8	12	16	20	24	28	32	36
11	0414	0453	0492	0531	0569						4	8	12	16	20	23	27	31	35
						0607	0645	0682	0719	0755	4	7	11	15	18	22	26	29	33
12	0792	0828	0864	0899	0934						3	7	11	14	18	21	25	28	32
						0969	1004	1038	1072	1106	3	7	10	14	17	20	24	27	31
13	1139	1173	1206	1239	1271						3	6	10	13	16	19	23	26	29
						1303	1335	1367	1399	1430	3	7	10	13	16	19	22	25	29
14	1461	1492	1523	1553	1584						3	6	9	12	15	19	22	25	28
						1614	1644	1673	1703	1732	3	6	9	12	14	17	20	23	26
15	1761	1790	1818	1847	1875						3	6	9	11	14	17	20	23	26
						1903	1931	1959	1987	2014	3	6	8	11	14	17	19	22	25
16	2041	2068	2095	2122	2148						3	6	8	11	14	16	19	22	24
						2175	2201	2227	2253	2279	3	5	8	10	13	16	18	21	23
17	2304	2330	2355	2380	2405						3	5	8	10	13	15	18	20	23
						2430	2455	2480	2504	2529	3	5	8	10	12	15	17	20	22
18	2553	2577	2601	2625	2648						2	5	7	9	12	14	17	19	21
						2672	2695	2718	2742	2765	2	4	7	9	11	14	16	18	21
19	2788	2810	2833	2856	2878						2	4	7	9	11	13	16	18	20
						2900	2923	2945	2967	2989	2	4	6	8	11	13	15	17	19
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8

LOGARITHMS

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9774	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9633	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	4	4

ANTILOGARITIIMS

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
·00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
·01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
·02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
·03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
·04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
·05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
·06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
·07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
·08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
·09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
·10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
·11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
·12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	3	3
·13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
·14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
·15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
·16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
·17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
·18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
·19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
·20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
·21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
·22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
·23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
·24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
·25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
·26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
·27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
·28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
·29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
·30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
·31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
·32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
·33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
·34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
·35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
·36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
·37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
·38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
·39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
·40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
·41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
·42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
·43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
·44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
·45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
·46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
·47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
·48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
·49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6

ANTILOGARITIIMS

	0	1	2	3	4	5	6	7	8	9	Mean Differences								
											1	2	3	4	5	6	7	8	9
·50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
·51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
·52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
·53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
·54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
·55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
·56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
·57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
·58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
·59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
·60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
·61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
·62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
·63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
·64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
·65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
·66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
·67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
·68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
·69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
·70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
·71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
·72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
·73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
·74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
·75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
·76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
·77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
·78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
·80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
·81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
·82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
·83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
·84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
·85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
·86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
·87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
·88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
·89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
·90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
·91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
·92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
·93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
·94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
·95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
·96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
·97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
·98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
·99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20

❐❐❐

Statistics and Computers for Animal and Veterinary Sciences

The present book has been well prepared to meet the requirements of the students of Animal and Veterinary Science, Animal Biotechnology and other related fields. The book will serve as a text book not only for students in Veterinary science but also for those who want to know "What statistics in all about" or who need to be familiar with at least the language and fundamental concepts of statistics. The book will serve well to build necessary background for those who will take more advanced courses in statistics including the specialized applications.

The salient features are:

- The book has been designed in accordance with the new VCI syllabus, 2016 (MSVE-2016).
- The book will be very useful for students of SAU's/ICAR institutes and those preparing for JRF/SRF/various competitive examinations.
- Each chapter of this book contains complete self explanatory theory and a fairly number of solved examples.
- Solved examples for each topic are given in an elegant and more interesting way to make the users understand them easily.
- Subject matter has been explained in a simple way that the students can easily understand and feel encouraged to solve questions themselves given in unsolved problems.

2018, 210 pages, figures, tables, 25cm

Rakesh Goel: *Incharge, Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, U.P. Pandit Deen Dayal Upadhyay Pashu Chikitsa Vigyan Vishwavidyalaya Evam Go Anusandhan Sansthan, Uttar Pradesh,Mathura, India*

nipa

NEW INDIA PUBLISHING AGENCY
101, Vikas Surya Plaza, CU Block, L.S.C. Market,
Pitam Pura, New Delhi - 110 034 (India)
Ph: +91-11-27 34 17 17, Fax: +91-11-27 34 16 16
Email: info@nipabooks.com
Website: www.nipabooks.com

978-93-86546-55-5

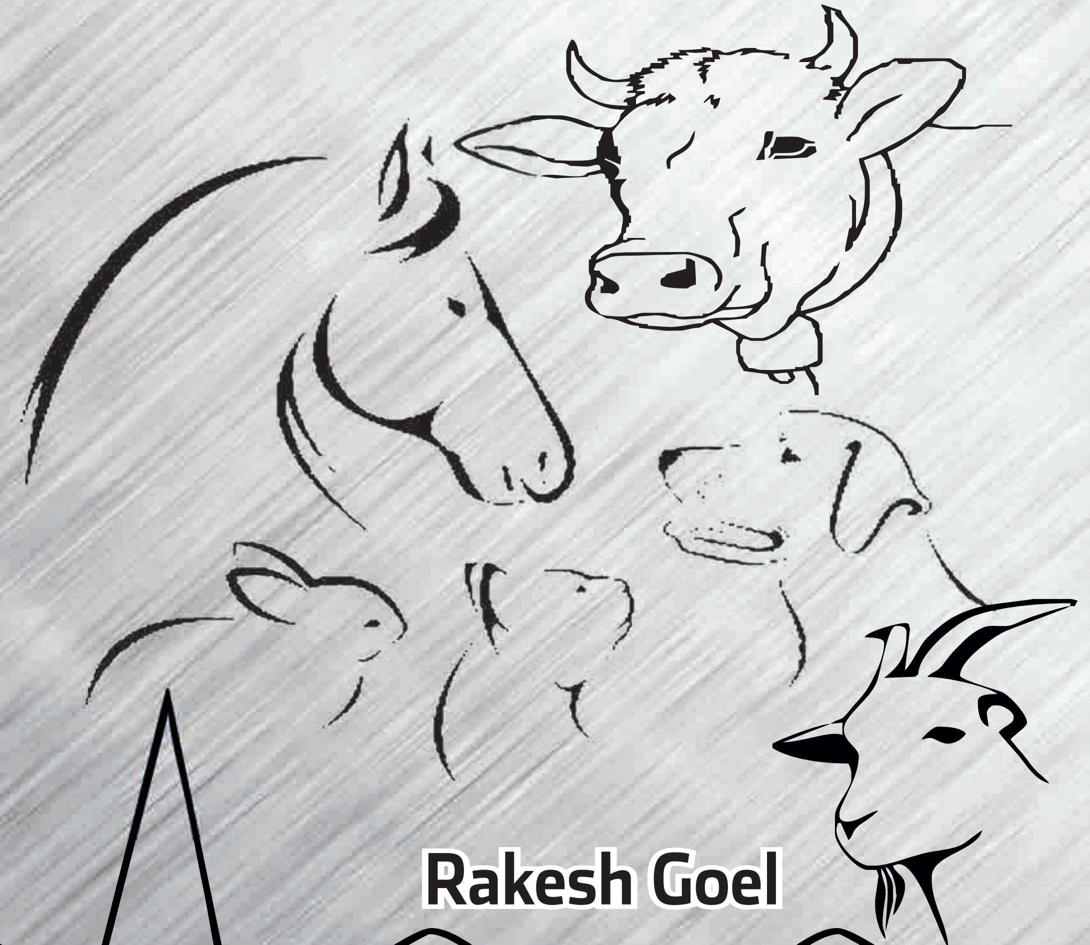
Statistics and Computers for Animal and Veterinary Sciences

Rakesh Goel

nipa

Statistics and Computers for Animal and Veterinary Sciences

Fundamentals and Applications



Rakesh Goel