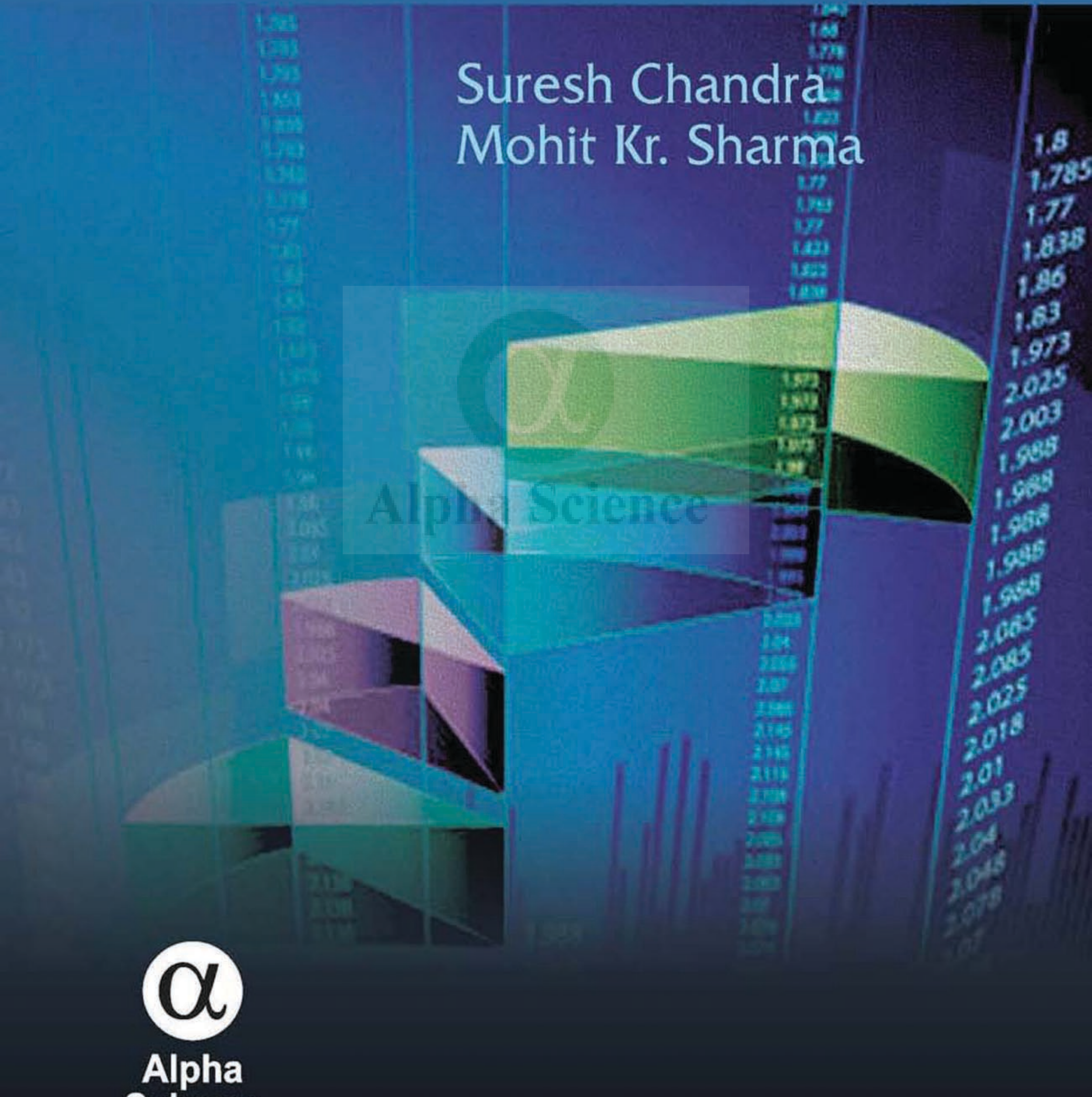
RESEARCH METHODOLOGY

Suresh Chandra
Mohit Kr. Sharma



Alpha Science

Research Methodology

Alpha Science

Research Methodology

Suresh Chandra
Mohit Kr. Sharma

Alpha Science International Ltd.
Oxford, U.K.

Research Methodology
266 pgs. | 17 figs. | 48 tbls.

Suresh Chandra
Department of Physics
Lovely Professional University
Phagwara, Punjab

Mohit Kr. Sharma
Research Student
School of Studies in Physics
Jiwaji University, Gwalior

ALPHA SCIENCE INTERNATIONAL LTD.
7200 The Quorum, Oxford Business Park North
Garsington Road, Oxford OX4 2JZ, U.K.

www.alphasci.com

Printed form the camera-ready copy provided by the Authors.

ISBN 978-1-84265-803-1

E-ISBN 978-1-78332-021-9

Printed in India

To my loving grand children, who represent the future.

Suresh Chandra

Preface

As a measure, to maintain standard of research, the University Grants Commission (UGC), New Delhi has made it compulsory to admit students for M. Phil. and Ph.D. programmes through entrance test. The pattern of this entrance test has also been provided in the form of NET (National Eligibility Test) exam which is being conducted by the UGC, New Delhi for non-science subjects and by the Council of Scientifc & Industrial Research (CSIR), New Delhi for science subjects. After qualifying such entrance test for M. Phil. or Ph.D. programme, students have to go through some courses and one of these courses is 'Research Methodology'. In this paper on Research Methodology, the students are given information: How to select topic for doing research; How to review literature; How to collect data (results of an investigation are data); How to analyze the data; How to interpolate between the data; How to estimate error in the calculations; How to calculate various statistical parameters for the results; How to present results; How to write a research paper and a thesis for the award of degree; What is a research journal; What is the difference between an Editor of a research journal and a Referee appointed for evaluation of research paper submitted the journal; What is Editorial Board of a research journal; Information about various techniques required for analyzing the data; Information about the hypotheses testing; Statistical test and other information pertaining to the Research Methodology. This paper also tells in brief about the application of computer in research. This information is essentially required for systematic investigation in research. This book describes the Research Methodology.

While preparing the manuscript, we consulted various books, research papers, handouts etc. We are grateful to the authors and publishers of those materials. For this book, we have been helped, advised and encouraged by our friends working in various institutes/universities in India as well as abroad, and by our friends in personal life. We are heartily thankful to all of them. We shall appreciate receiving comments/suggestions, which can be sent at the following emails:

suresh492000@yahoo.co.in   mohitkumarsharma32@yahoo.in

Thanks are due to Prof. Dr. W.H. Kegel, Prof. S.M. Karuppayil and Dr. Niti Kant for encouragement. Suresh Chandra is specially thankful to his wife Mrs. Purnima Sharma for her valuable cooperation in his life and for sharing a major part of responsibility of family affairs. We are deeply grateful to our family members who have always been a source of inspiration and happiness for us. We are also thankful to Prof. C. Bertout, Editor-in-Chief, A&A for permitting us to include rearch paper printed the end of chapter I. Last, but not the least, we are highly thankful to the publisher of this book.

Suresh Chandra
Mohit Kr. Sharma

Contents

I. Introduction

An institution is known by quality of research work going on there, and the quality of research work is determined on the basis of appreciation given by the peer groups. A quality research work can obviously be carried out when the institution has highly qualified and dedicated faculty. Though buildings, laboratories and other infrastructure are essentially required for performing the research work, but the main contribution for recognition of an institution is assigned to the faculty working there. It is not out of context to mention that the cost of the apparatus used by Sir C.V. Raman was meager, but when it was combined with the mind set of Raman, the result came out in the form of Noble Prize in Physics to India in 1930. To commemorate the discovery of Raman effect, India celebrates Science Day on 28 February each year.

India got two more Noble Prizes: One in Literature (Guru Rabindra Nath Tagore in 1913) and second in Peace (Mother Teresa in 1979). Moreover, we can proudly mention that four personalities born in India have been honoured with Noble Prize. They are (i) Har Gobind Khorana (Medicine in 1968), (ii) Subrahmanyan Chandrasekhar (Physics in 1983), (iii) Amarteya Sen (Economics in 1998), and (iv) Venkataraman Ramakrishnan (Chemistry in 2009). It is also interesting to know that Subrahmanyan Chandrasekhar was nephew of Sir C.V. Raman. That is, two persons from a common family got Noble Prize in Physics independently.

Efforts from the Government of India for development of science and technology are tremendous and continuously going on for several decades and many good research institutions in various disciplines have been established. These institutions are performing well and the work done in India is being recognized and appreciated in various forums at the international as well as national levels. For doing good research work, one has to proceed step by step in a disciplined manner.

For the success in any field, discipline is an essential part. Generally, the word 'discipline' is mentioned in the context of army. The meaning of this word may be taken as the 'systematic manner'. Hence, this is also applicable to our day-to-day life as well as to all sorts of studies, including research in a specific field of interest. Research is being done in various disciplines, *e.g.,* science, technology, business and management, social sciences, languages, etc. in India as well as abroad. The purpose of research in business and management is to enhance the profit while in science and technology is to invent new techniques or to divulge unknown facts.

Quite often, the research is done with the aim of getting a degree, called Ph.D. (Doctor of Philosophy). The main purpose of doing research is to enhance knowledge in the concerned field, and in turn that should be beneficial for the society. The mode of benefit to the society may be explicit or implicit. Many research works have indirect link for the benefit of society.

Suppose a person wants to do research, one has to make a proper plan. How to proceed for doing a successful research work will be an attempt for discussion in the present chapter.

1. Joining research

For doing research, one has to join an institution and the following two points come first for the consideration:

(i) Selection of an institution

(ii) Selection of a guide

1.1 Selection of an institution

The word 'institution' used here includes the research institutes and universities. A university may have affiliated colleges besides a campus. In every field, a number of research institutes and universities are offering research programs in India as well as abroad. For doing research, one obviously would like to choose the best available institution. But, limited number of seats and entrance exam are the main hurdles which one has to surmount. Students in their mind have different priorities about the institutions. For example, in India, IIT's, IIM's and some well recognized institutions are considered at the top priorities. These priorities are because of their recognitions earned by them on the basis of quality of research work going on there. A large number of research institutions conduct their own entrance exam for selection of students. Central bodies, such as UGC, CSIR, etc, are conducting the NET (National Eligibility Test). Some State Governments in India are also conducting similar kind of test, called the State Level Eligibility Test, abbreviated as SET or SLET. Hence, for getting an institution, one has to qualify an entrance test. Moreover, seats in these institutions are limited. In the event of not getting seat in an institution at high priority, one has to look for an institutions at low priority.

Earlier, in many universities, there was no provision for such entrance test. But, now the UGC, New Delhi has made it mandatory. Even for the admission for M.Phil (Master of Philosophy) programme, one has to go through an entrance test. However, when one wants to go for Ph.D. program after completing M.Phil. programme, in the same institution, there is no need for entrance test. However, other institutions may ask to write again the entrance exam for joining Ph.D. programme. Let us wish each one gets an institution of his/her choice, at high priority or at low priority.

1.2 Selection of a guide

After joining an institution for research work for doing M.Phil. or Ph.D., as per guidelines of the UGC, New Delhi, the selected candidates go through a set of courses. When a candidate goes for Ph.D. after completion of M.Phil., in some institutions, the course work may not be required, as it is already done during M.Phil. programme. This set of courses essentially provides information about the procedure of research work and about some basic training for doing research. About this we shall discuss after a while. After successful completion of the course, one needs a guide for supervision of the research work.

In an institution, there may be several scientists/teachers guiding students for their research work. Then the question arises how to select one of them. While selecting a guide, the first consideration should to do research work in the field of latest developments in the area. Again there may be priorities for selection of a guide. Obviously, a well reputed guide may have limited number of vacancies for new students to join. One more point, one should consider that the guide should be cooperative and of encouraging and supportive nature. Many times it may be true that getting a good guide is a big success towards doing research. On the other side, in some institutions, guides are allocated by an appropriate committee meant for it.

2. Basic courses

Depending on the nature of the research work, an institution designs and delivers a course program to the research students admitted there. As per directions of the UGC, New Delhi, in this program, one paper, called 'Research Methodology' is mandatory. The course content of Research Methodology can be different, depending on the discipline. For example, the content for engineering and science may be different from that for for the applied medical sciences and life sciences, which may again be different from that for commerce and management. Broadly, it can be classified into two groups of Ph.D. program : (i) where knowledge of mathematics is used and (ii) where use of mathematics is up to the level of basic statistics. Moreover, some content in all the disciplines will be common.

Nowadays, the use of computer is almost mandatory for each and every study. One has to prepare manuscript with the help of computer. While preparing a manuscript, one requires to type text, to prepare graphs, diagrams, charts, etc,, to write formulas. Moreover, presentation of the manuscript can be made very effective with the help of some specific softwares. After obtaining the results/data in a study, their analysis may be done with the help of a computer software. So, it becomes essential to teach the students about computer and about preparation of manuscript and dealing with data.

In the paper of Research Methodology, the students, are given information about various aspects pertaining to the research program.

3. Basic steps for doing research

Here, we shall discuss about some steps to be taken while doing research. These are just guidelines and there may be more factors as well.

3.1 Literature survey

Broad area of research field of a research student depends on the research field of the guide, assigned/chosen for doing the research work. After explaining broad features of research work going on in the group, the guide generally asks the student for doing survey of literature in the field. The survey provides the knowledge/information about the latest status of research work published by peer groups in the field.

In early days, one had to visit various libraries in different institutions for the survey of literature. For that one had to spend lot of time and money. Nowadays, the internet has made this task very convenient and economic. Most of the information can be accessed with the help of the internet. Moreover, the internet provides extensive and latest information in the field of interest. Nevertheless, some journals permit up to the extent of Abstract of the research paper and ask for payment when one is interested in the full text of the paper. On the other side, it is interesting to note that many books are also available on the internet.

In the literature survey, the very first step is to gather the latest as well as old information about the relevant research work published by the peer groups all over the world. After going through the literature, one can decide about the specific topic for his/her research work. The topic is also discussed with the research guide. As the guide, in general, has vast experience of the research field, he/she can help the student in finalizing the research topic for investigation.

3.2 Investigation

After finalizing the research topic, one can prepare his/her plan of research work and submit through proper channel to the authorities of the institution for approval. This procedure for approval may take a time of some months. After submission of proposal, with the hope that the proposal would be approved, one can start the investigation pertaining to research work. The nature of investigation may be experimental, theoretical or collection of information from various sources. Hence, in the investigation, one finally has data which can are to be analyzed.

In the experimental study, one performs various experiments. In case, the required experiments are not available in his/her institution, one takes help from other institutions. The theoretical investigation, in general, depends on the computer and relevant softwares written by various research groups. The success of research work depends on the depth of the investigation.

3.3 Analysis of data

In an investigation, finally data are obtained. These data may be i the form of output of experimental work, theoretical calculations, or survey in the society. After obtaining data from investigation, the next step is to analyze them, so that some meaningful conclusions could be drawn. The analysis can be done by plotting graphs, by performing some tests with the help of computer. In the present book, we shall discuss about interpolation, error analysis, chi-square test.

3.4 Preparation of manuscript

After analyzing the data, when one gets encouraging results, he/she would like to publish them for the knowledge and benefit of others as well. These results may also be used as valuable input information for the investigation of others. Hence, one has to prepare manuscript either for publication of a research paper in a research journal or for submission of a thesis for the award of Ph.D./M.Phil. degree by the institution. In the preparation of manuscript, one has to write text pertaining to the research work, to prepare tables and graphs/diagrams. The manuscript may be written in the form of various sections in a research paper and in the form of various chapters in a thesis. In the 'introduction' section/chapter, one has to write about the background of the work and goal of the investigation.

3.5 Publication of research work

The manuscript prepared in the form of a research paper is submitted to an Editor or Editor-in-chief of a research journal for its publication. One always desires to publish his/her paper in a journal having high impact-factor. However, publication charges in some journals may come as a hurdle for publication, in particular for those who have no funds to pay as publication charges. The paper submitted to a journal is sent by the editorial office to one or two referees for their comments on the content of the paper. Nowadays, many journals have prepared specific format according to which a manuscript is prepared and submitted.

The referees are generally renowned/distinguished persons working in the concerned research field. After study of the manuscript, the referees may recommend as the following:

- Acceptance of the paper
- Revision of the paper
- Rejection of the paper

In the event of acceptance of the paper, the corresponding author is asked for sending the soft-copies of the manuscript and figures, so that the work and manuscript could be modified and published as per format of the journal.

In the event of the revision of the paper, the reports received from the anonymous referees are sent to the corresponding author. The authors then react on the reports. They may agree to the suggestions/comments received from the referees and then revise the paper. The revised manuscript is then submitted to the Editor/Editor-in-chief which in turn is sent to the referee again for review. In case, the referees are satisfied with the revision of the manuscript, they can recommend for accepting the paper. When the revision is not to the satisfaction of the referees, the referees may ask for further revision or may recommend for rejection of the paper.

In the event of rejection of the paper, the Editor/Editor-in-chief informs to the corresponding author about it, and the authors can submit the paper to another research journal in the field.

Example of a research paper

Photocopy of a research paper published by one of the authors of this book is included here as an example. The paper is printed at the end of this chapter. This paper is published in the research journal named 'Astronomy & Astrophysics'. For printing of this paper, proper permission is taken form the Editor-in-Chief, Prof. C. Bertout of the journal vide letter dated 30 August 2011. This journal has Impact factor 4.410, which is quite high. Impact factor of a journal is a measure of standard of the journal. This journal is published in two parts: one part publishes regular research papers and Short Communications whereas the other part publishes Letters. The purpose of Letters is to disseminate scientific findings at the earliest possible. Moreover, the work published in Letters is considered to be more important from the scientific point of view. Each part of the journal, Astronomy & Astrophysics, has one Editor-in-Chief. This journal has a Board of Directors. (Some journals have only Editor and no Board of Directors.) All these positions have a well defined tenure. At present Editors-in-Chief are as the following.

Editor-in-Chief: C. Bertout

Letters Editor-in-Chief: M. Walmsley

From website of the journal, it can be found that this journal was a result of merging of the following journals. These journals agreed to bring out a common journal Astronomy & Astrophysics.

Annales d'Astrophysique

Arkiv for Astronomi

Bulletin of the Astronomical Institutes of the Netherlands

Bulletin Astronomique

Journal des Observateurs

Zeitschrift für Astrophysik

Bulletin of the Astronomical Institutes of Czechoslovakia

The member countries and present Board of Directors of Astronomy & Astrophysics are as the following. (Information received from website of the journal on dated 1 September 2011.)

Member Countries	Board of Directors	Member Countries	Board of Directors
Argentina	J. Zorec	Germany	N. Langer
Austria	M. Breger	Greece	N. Prantzos
Belgium	C. Sterken	Hungry	L.L. Kiss
Brazil	W.J. Maciel	Italy	G. di Cocco
Bulgaria	K. Panov	Lithunia	A. Kucinskas
Chile	M. Rubio	Netherlands	J. Lub
Croatia	V. Ruzdjak	Poland	B. Rudak
Czech Republic	M. Wolf	Portugal	J.F. Gameiro
Denmark	B. Nordstrom	Slovak Republic	A. Skopal
ESO*	B. Leibundgut	Spain	E. Perez
Estonia	L. Leedjarv	Sweden	J. Solleman
Finland	J. Harju	Switzerland	G. Meynet
France	F. Genova		

*ESO is the European Organisation for Astronomical Research in the Southern hemisphere.

Notice the following about the paper.

Title of paper: Absorption against the cosmic 2.7 K background

Names of the authors: S. Chandra and W.H. Kegel

Affiliation of the authors:

School of Physical Sciences, S.R.T.M. University, Nanded 431606, India

Institut für Theoretische Physik, J.W. Goethe-Universität, 60054 Frankfurt am Main, Germany

Key words: molecular processes-ISM:molecules-Cosmology:cosmic microwave background-radio lines: ISM

Abstract: Giving information about the paper in brief.

Introduction:

Details of the investigation

Results:

Discussion:

Acknowledgments:

References: Addresses where the referred papers are published.

Activities

(i) Using the internet, find out some research journals published in the field of your interest. For five journals, write down the following information:

A. Name of Editor/Editor-in-Chief

B. Frequency of publication

C. Editorial board, if applicable

D. Name of publishing agency

(ii) Down load three research papers and write down about the headings of the papers.

4. Problems and questions

1. What are major steps taken for joining research work in an institution.

2. Discuss the relevance of basic courses in doing research.

3. Write short notes on the following:

   (i) Literature survey

   (ii) Investigation during research

   (iii) Analysis of data

   (iv) Preparation of manuscript

   (v) Role of Editor/Editor-in-chief of a research journal.

   (vi) Referee of a research journal and his/her role.

A&A 367, 995–999 (2001)
DOI: 10.1051/0004-6361:20000499

Astronomy & Astrophysics

Absorption against the cosmic 2.7 K background

S. Chandra[1,2] and W. H. Kegel[2]

[1] School of Physical Sciences, S.R.T.M. University, Nanded 431 606, India
[2] Institut für Theoretische Physik, J.W. Goethe-Universität, 60054 Frankfurt am Main, Germany

Received 25 February 2000 / Accepted 12 December 2000

Abstract. Two interstellar lines, $1_{10} \rightarrow 1_{11}$ of formaldehyde at 4.831 GHz, and $2_{20} \rightarrow 2_{11}$ of cyclopropenylidene at 21.590 GHz, have so far been observed in absorption against the cosmic 2.7 K background. Observation of an interstellar line in absorption against the cosmic 2.7 K background is an unusual phenomenon, and can only be possible under rather peculiar conditions developed in the molecule, generating the line. We predict that two more lines, $3_{30} \rightarrow 3_{21}$ at 27.100 GHz, and $3_{31} \rightarrow 3_{22}$ at 59.550 GHz of cyclopropenylidene, and three lines, $2_{20} \rightarrow 2_{11}$ at 15.600 GHz, $3_{30} \rightarrow 3_{21}$ at 23.100 GHz, and $3_{31} \rightarrow 3_{22}$ at 39.700 GHz of ethylene oxide, may show absorption against the cosmic 2.7 K background. We speculate that such peculiar conditions are characteristic for *b*-type asymmetrical top molecules. – The observation of these lines may be used to place upper bounds to the density in the absorbing region.

Key words. molecular processes – ISM: molecules – Cosmology: cosmic microwave background – radio lines: ISM

1. Introduction

Observation of an interstellar line in absorption against the cosmic 2.7 K background is an unusual phenomenon. The intensity, I_ν, of a line generated in an interstellar cloud, with homogeneous excitation conditions, is given by

$$I_\nu - I_{\nu,\mathrm{bg}} = (S_\nu - I_{\nu,\mathrm{bg}})(1 - \mathrm{e}^{-\tau_\nu})$$

where $I_{\nu,\mathrm{bg}}$ is the intensity of the continuum against which the line is observed, τ_ν the optical depth of the line, and S_ν the source function, which is the Planck's function at the excitation temperature T_{ex}, i.e., $S_\nu = B_\nu(T_{\mathrm{ex}})$. For positive optical depth (which is a usual situation for most of the lines), a line is observed in absorption when $S_\nu < I_{\nu,\mathrm{bg}}$. The lower limit of $I_{\nu,\mathrm{bg}}$ is given by the cosmic 2.7 K background. Observation of an interstellar line in absorption against the cosmic 2.7 K background, obviously, implies $T_{\mathrm{ex}} < 2.7$ K, which requires rather peculiar physical conditions in the molecule, generating the line.

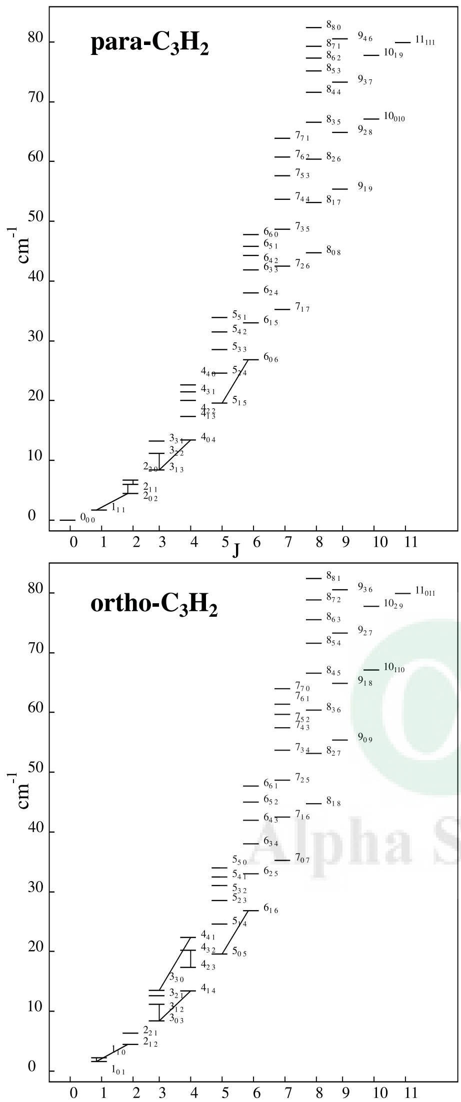
Up to now, only two lines have been reported in absorption against the cosmic 2.7 K background. The first one is the $1_{10} \rightarrow 1_{11}$ transition of formaldehyde at 4.831 GHz, which was found in absorption in several directions (Palmer et al. 1969). However, in some cases, it has been seen in emission, and even as a maser line (Foster et al. 1980; Whiteoak & Gardener 1983). The second line found in absorption against the cosmic 2.7 K background, in a large number of cosmic objects, is the $2_{20} \rightarrow 2_{11}$ transition of cyclopropenylidene at 21.590 GHz (Madden et al. 1989). Cox et al. (1987), however, reported the observation of this line in emission in the Planetary Nebula NGC 7027.

Under interstellar conditions, the relative occupation of the levels in a given molecule is controlled by the competition between collisional and radiative transitions, between the energy levels in the molecule, and in general, it cannot be described by the Boltzmann distribution law. There are, however, two simple limiting cases: (i) For very high densities in the line forming region, the collisional transitions dominate the radiative ones, and thus, one gets a Boltzmann distribution corresponding to the local kinetic temperature, T_{kin}. (ii) When the density in the region is so low that the collisional transitions may be neglected in comparison to the radiative ones, and the external radiation field is just the microwave background, one gets the Boltzmann distribution corresponding to 2.7 K. (This, in fact, is one way to measure the temperature of the cosmic background radiation, Crane et al. 1990.) For intermediate densities, one finds, in the two level approximation, 2.7 K $\leq T_{\mathrm{ex}} \leq T_{\mathrm{kin}}$. This shows that in order to find out physical conditions, under which $T_{\mathrm{ex}} < 2.7$ K may occur, one has to account for a multilevel system of the molecule under investigation.

2. Formulation of the problem

In the present work, we have investigated two molecules, cyclopropenylidene (C_3H_2), and ethylene oxide (C_2H_4O). Cyclopropenylidene was identified in the interstellar space by Thaddeus et al. (1985) and has been found in many objects (Madden et al. 1989). Out of the thirteen lines

Send offprint requests to: W. H. Kegel,
e-mail: kegel@astro.uni-frankfurt.de



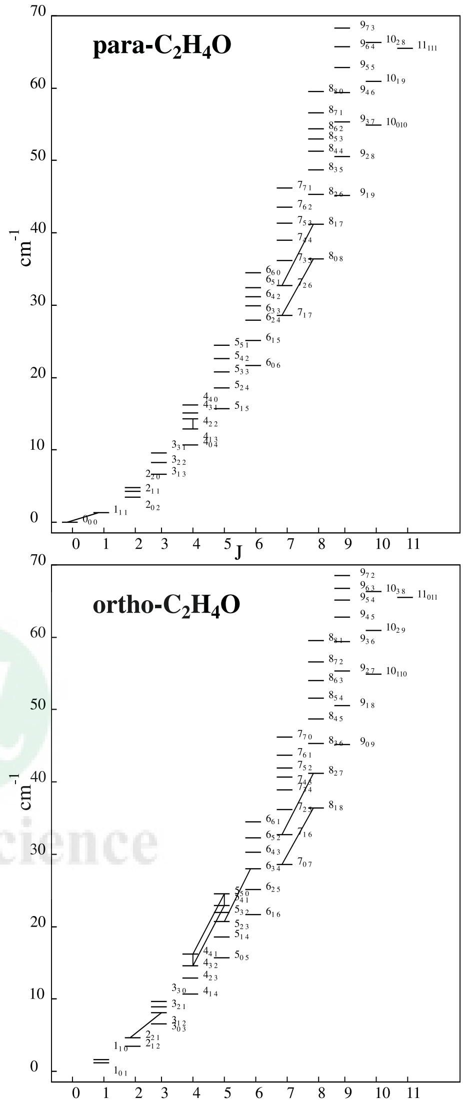


Fig. 1. Rotational energy levels in the ground vibrational state of para- and ortho-cyclopropenylidene (C_3H_2). The transitions shown correspond to the observed lines in the interstellar clouds

Fig. 2. Rotational energy levels in the ground vibrational state of para- and ortho-ethylene oxide (C_2H_4O). The transitions shown correspond to the observed emission lines in Sgr B2N

observed (see, Fig. 1) only the $2_{20} \rightarrow 2_{11}$ transition has been found in absorption. Ethylene oxide has been observed recently (Dickens et al. 1997), in Sgr B2N, through its twelve emission lines (see, Fig. 2). Both molecules are *b*-type asymmetric oblate rotors, having two distinct species, ortho and para.

With the aim to determine the physical conditions required for occurring the anomalous absorption against the cosmic 2.7 K background, we performed NLTE radiative transfer calculations for cyclopropenylidene (C_3H_2), and ethylene oxide (C_2H_4O), using a large velocity gradient (LVG) code (Cox et al. 1987; Rausch et al. 1996;

Table 1. Value of the minimum excitation temperature (Col. 4) achieved for the lines, showing absorption against the cosmic 2.7 K background, around the molecular hydrogen density n_{H_2} (Col. 5) at the kinetic temperature of 30 K. Columns 6 and 7 give the radiative life time of the upper and lower levels, respectively, of the line

Specie	Transition	Frequency (GHz)	T_{ex} (K)	n_{H_2} (cm^{-3})	Radiative Life Time (s)	
					Upper level	Lower level
para-C_3H_2	$2_{20} \rightarrow 2_{11}$	21.590	1.1	5 10^4	1.79 10^4	3.61 10^5
ortho-C_3H_2	$3_{30} \rightarrow 3_{21}$	27.100	1.6	5 10^4	3.76 10^3	1.56 10^4
para-C_3H_2	$3_{31} \rightarrow 3_{22}$	59.550	2.1	2 10^4 (at large γ)	3.06 10^3	7.94 10^3
para-C_2H_4O	$2_{20} \rightarrow 2_{11}$	15.600	0.8	3 10^4	1.39 10^5	6.12 10^6
ortho-C_2H_4O	$3_{30} \rightarrow 3_{21}$	23.100	1.1	1 10^5	2.74 10^4	1.25 10^5
para-C_2H_4O	$3_{31} \rightarrow 3_{22}$	39.700	1.5	3 10^4	2.43 10^4	5.86 10^4

de Jong et al. 1975; Goldreich & Kwan 1974), where the physical model is that of a homogeneous collapsing cloud. Owing to its simplicity, the present model allows to vary the physical parameters over wide ranges, with a moderate numerical efforts. The molecular data required as input parameters are: (i) Einstein coefficients for the various radiative transitions between the rotational energy levels accounted for (Figs. 1 and 2), and (ii) the rate coefficients for collisional transitions between the levels due to collisions with H_2 molecules. The Einstein A-coefficients are taken from Sharma & Chandra (1996) for cyclopropenylidene, whereas for ethylene oxide are calculated by using the molecular and distortional constants derived by Pan et al. (1998); the collisional rate coefficients for the transitions in cyclopropenylidene are taken from Avery & Green (1989), and Chandra & Kegel (2000), whereas for the downward transitions $J'k'_ak'_c \rightarrow Jk_ak_c$ in ethylene oxide at the kinetic temperature of 30 K, are approximated by the relation

$$C(J'k'_ak'_c \rightarrow Jk_ak_c) = 1\ 10^{-11}/(2J'+1).$$

Since this relation has no particular selections, it cannot from its own generate any NLTE situation. The rate coefficient for the corresponding upward transitions $Jk_ak_c \rightarrow J'k'_ak'_c$ has been calculated with the help of the detailed equilibrium equation.

As a background radiation field, we accounted only for the 2.7 K microwave radiation. The value of the molecular hydrogen density, n_{H_2}, which determines the collisional rates, and the quantity $\gamma \equiv n_{mol}/(dv_r/dr)$ (where n_{mol} is the density of the molecule and dv_r/dr the velocity gradient in the region), which corresponds to the column density of the molecule for the distance over which the Doppler shift is equal to the thermal line width, and determines the optical thickness in the various lines, have been varied over the ranges shown in Fig. 3. For cyclopropenylidene, the calculations were performed for T_{kin} = 10, 20, 30, 60, and 120 K. We found anomalous absorption, i.e., $T_{ex} < 2.7$ K, in both the molecules for three lines, $2_{20} \rightarrow 2_{11}$, $3_{30} \rightarrow 3_{21}$, and $3_{31} \rightarrow 3_{22}$ (Fig. 3, Table 1).

3. Results

Since the line $2_{20} \rightarrow 2_{11}$ of cyclopropenylidene has already been observed in anomalous absorption, let us first, discuss, in brief, the results for this transition. In the present investigation, the basic excitation is evidently caused by collisions. Since the anomalous absorption is observed for a line connecting two excited levels, one would expect, at low densities, the effect to increase with the increase of n_{H_2}, and in fact this is what our numerical results show. The value of T_{ex} initially decreases with increasing n_{H_2}, reaches a minimum value and then increases. For T_{kin} = 10 K, the minimum value of T_{ex} is 1.3 K, which occurs around $n_{H_2} \approx 10^5$ cm^{-3}. For $n_{H_2} \geq 10^6$ cm^{-3}, we found $T_{ex} > 2.7$ K. Since the rate coefficients for collisional excitations increase with the kinetic temperature, the effect is expected to increase with the kinetic temperature. Our results are found in agreement with this expectation. At $T_{kin} = 60$ K, we find a minimum excitation temperature of 1.1 K occurring around $n_{H_2} = 5\ 10^4$ cm^{-3}. In the optically thin limit, i.e., when all lines are optically thin, the relative occupation numbers are independent of the column density. In this case, the optical depth, at a given molecular hydrogen density, increases linearly with the column density of the molecule. According to our results, optical thickness effects become important when $\gamma \geq 10^{-6}$ $cm^{-3}(km\,s)^{-1}$ pc. For larger values of γ, the excitation temperature of the $2_{20} \rightarrow 2_{11}$ transition is found to increase with the increase of the column density of the molecule.

As an illustration, Fig. 3 gives the intensities calculated for T_{kin} = 30 K, where iso-lines of $(I_\nu - I_{\nu,bg})/B_\nu(T_{kin})$ have been plotted. The iso-line for the value -10^{-2}, for example, corresponds to the absorption of 0.3 K in brightness temperature.

We find qualitatively similar results for the $3_{30} \rightarrow 3_{21}$ and the $3_{31} \rightarrow 3_{22}$ transitions. At T_{kin} = 30 K, the effect seen in the intensity (Fig. 3) is an order of magnitude smaller for the $3_{30} \rightarrow 3_{21}$ transition as compared to the $2_{20} \rightarrow 2_{11}$ transition. The anomalous absorption in the $3_{31} \rightarrow 3_{22}$ transition is about two orders of magnitude

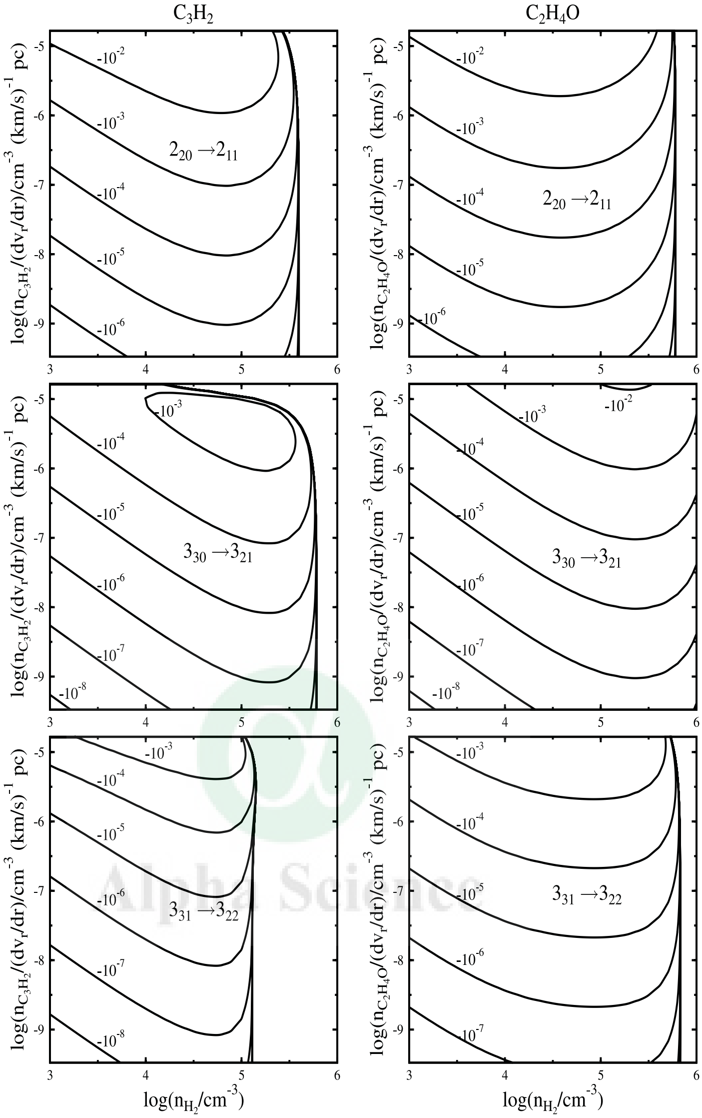


Fig. 3. Iso-lines for $(I_\nu - I_{\nu,\mathrm{bg}})/B_\nu(T_{\mathrm{kin}})$ for the spectral lines showing absorption against the 2.7 K background, computed for $T_{\mathrm{kin}} = 30$ K

weaker than in the $2_{20} \rightarrow 2_{11}$ transition and its detection will be difficult.

It appears worthwhile to discuss the numerical results in a qualitative manner in order to identify the main physical effect leading to anomalous absorption in the case studied here. We concentrate on the $2_{20} \rightarrow 2_{11}$ transition for which the effect is strongest. – At first we note that the pump mechanism is distinct from that leading to anomalous absorption in H_2CO. In the latter case anomalous absorption is observed for the line connecting the two lowest energy levels of ortho-H_2CO. To have an overpopulation of the 1_{11} state there must be an effective way to transfer molecules in the 1_{10} state to the 1_{11} state. According to Townes & Cheung (1969) this is achieved by the selectivity of the collision cross-sections, which favour the excitation of the 2_{12} state over that of the 2_{11} state. Subsequent radiative decay leads to an overpopulation of the 1_{11} state. – In the case studied here, we found anomalous absorption for transitions connecting two excited levels for which the lifetime against radiative decay of the lower level is substantially larger than that of the upper level. In the case of the $2_{20} \rightarrow 2_{11}$ transition the ratio of the lifetimes is

20 (Table 1). Anomalous absorption occurs in a density regime in which the lifetime of the upper level against collisional transitions is larger than that against radiative transitions. We further note that in the approximation used by Avery & Green (1989), the cross-section for collisional excitation of the 2_{11} level as well as of the 3_{22} level from the ground state 0_{00} is zero. Therefore, the sum of the rates of collisional excitation from the two lowest levels (0_{00} and 1_{11}) is smaller for the 2_{11} level than for the 2_{20} level. This implies that the collisions by themselves have the tendency to favour an overpopulation of the upper level rather than of the lower one. (In fact for very large densities our numerical results indicate population inversion.) These facts show that anomalous absorption in the present case is caused essentially by the differences of the radiative lifetimes of the upper and the lower level and not – as in the case of H_2CO – by the selectivity of the collisions. The main role of the collisions is to give a general excitation of the molecules, while the unequal distribution over the different levels is mainly caused by the radiative transitions.

Avery & Green (1989) investigated cyclopropenylidene accounting for 16 and 17 rotational energy levels for ortho- and para-C_3H_2, respectively, and found anomalous absorption, besides the transition $2_{20} \rightarrow 2_{11}$, for $3_{30} \rightarrow 3_{21}$ and $3_{21} \rightarrow 3_{12}$ transitions. Our calculations, accounting for 47 and 48 energy levels for ortho- and para-C_3H_2, respectively, (see, Fig. 1), however, did not find anomalous absorption for the $3_{21} \rightarrow 3_{12}$ transition. We note in passing that we found population inversion for the $1_{11} \rightarrow 0_{00}$ transition for n_{H_2} in the range from 10^5 cm^{-3} to 10^6 cm^{-3}, and for the $3_{13} \rightarrow 2_{20}$ transition in the range from 10^5 cm^{-3} to 3 10^5 cm^{-3}. The inversion is, however, so weak that it would not be recognized from the line profiles.

The results obtained for ethylene oxide are qualitatively very similar to those for cyclopropenylidene (see, Fig. 3), in particular we found anomalous absorption for the corresponding transitions. We relate this to the similarity of the energy level diagrams (Figs. 1 and 2) and the relative values for the Einstein A-coefficients for both the molecules. According to the arguments given above this implies that the occurrence of anomalous absorption against the cosmic 2.7 K background in these two systems is primarily related to peculiarities of the energy level diagram and the relative values of the Einstein A-coefficients, and to a lower degree to the collisional rate coefficients. In Table 1, we have given the minimum value of the excitation temperature $T_{\rm ex}$, achieved for the line, around the molecular hydrogen density given in Col. 5 of the table.

A necessary condition for the mechanism discussed above to work, is that the radiative life-time of the upper level of the line must be smaller than that of the lower one. The larger the ratio of the life-times of the lower to the upper levels, the larger the absorption against the cosmic 2.7 K background. In case of anomalous absorption, the ratio of the population densities of the upper to the lower levels of the line satisfies the condition, $(n_{\rm u}/n_{\rm l})_{T_{\rm ex}} < (n_{\rm u}/n_{\rm l})_{2.7}$. As can be seen from Fig. 3, there is a rather sharp upper limit in density above which anomalous absorption does not occur. Thus the observation of anomalous absorption of the lines discussed is a rather direct way to place an upper bound to the density. At the critical density the collisional life-times of the levels become comparable to the radiative ones.

Based on the similarity of our numerical results for cyclopropenylidene and ethylene oxide, and based on the qualitative arguments given above, we expect that the occurance of anomalous absorption against the cosmic 2.7 K background, in particular for the $2_{20} \rightarrow 2_{11}$ transition may be a characteristic of b-type asymmetric top molecules.

Acknowledgements. This work was done during the visit of S. C. in Frankfurt under the scientific exchange program between the INSA, New Delhi (India) and the DFG, Bonn (Germany).

References

Avery, L. W., & Green, S. 1989, ApJ, 337, 306
Chandra, S., & Kegel, W. H. 2000, A&AS, 142, 113
Cox, P., Güsten, R., & Henkel, C. 1987, A&A, 181, L19
Crane, P., et al. 1990, in The Quest for Fundamental Constants in Cosmology, ed. J. Audouze, & J. Than Van (Éditions Frontières)
de Jong, T., Chu, S.-I., & Dalgarno, A. 1975, ApJ, 199, 69
Dickens, J. E., Irvine, W. M., Ohishi, M., et al. 1997, ApJ, 489, 753
Foster, J. R., Goss, W. M., Wilson, T. L., Downes, D., & Dickel, H. R. 1989, A&A, 84, L1
Goldreich, P., & Kwan, J. 1974, ApJ, 189, 441
Madden, S. C., Irvine, W. M., Matthews, H. E., et al. 1989, AJ, 97, 1403
Palmer, P., Zuckerman, B., Buhl, D., & Snyder, L. E. 1969, ApJ, 156, L147
Pan, J., Albert, S., Sastry, K. V. L. N., et al. 1998, ApJ, 499, 517
Rausch, E., Kegel, W. H., Tsuji, T., & Piehler, G. 1996, A&A, 315, 533
Sharma, A. K., & Chandra, S. 1996, J. A&A, 17, 41
Thaddeus, P., Vrtilek, J. M., & Gottlieb, C. A. 1985, ApJ, 299, L63
Townes, C. H., & Cheung, A. C. 1969, ApJ, 157, L103
Whiteoak, J. B., & Gardner, F. F. 1983, MNRAS, 205, 27p

II. Error Analysis

The error analysis is a study and evaluation of uncertainty in measurement of a physical quantity. It is not difficult to understand that no measurement, even made carefully, can be completely free from uncertainty. In science, the term 'error' is used in the sense of uncertainty in a measurement. As such errors are not mistakes. One cannot avoid them by being very careful. The best one can do is to (i) find reliable estimates of their size, and (ii) use experimental designs and procedures that keep them as small as possible. A better way to estimate uncertainty is to make multiple measurements of the same quantity and to analyze the data set with the help of statistical methods.

1. Random and systematic errors

Though a better way to estimate uncertainty in a measurement is to make multiple measurements of the same quantity, but not all types of experimental uncertainties can be assessed with the help of repeated measurements. For this reason, uncertainties are classified into two groups: (i) random uncertainties, which can be treated statistically, and (ii) systematic uncertainties, which cannot be treated statistically. Random uncertainties can be estimated by repeating the measurements whereas the systematic uncertainties cannot be estimated.

In order to understand the random and systematic uncertainties, let us consider an example of simple pendulum where we measure the time period (time for one oscillation) of a steadily moving pendulum. One source of error will be our reaction time in starting and stopping the stop-watch. If our reaction times are always exactly the same, these two delays would cancel one another. In practice, our reaction time may vary each time. We may delay more in starting the stop-watch, and so underestimate the time of one oscillation, or we may delay more in stopping the stop-watch, and so overestimate the time of one oscillation. Since either possibility is equally likely, both the size and the sign of the effect are random. If we repeat the measurement several times, our variable, the reaction time will show a variation in the measured periods. By looking at the spread in the measured periods, we can get a reliable estimate of this kind of error.

On the other side, when our stop-watch is running consistently slow, then all our times will be underestimated, and no amount of repetition (with the same watch) will estimate this kind of error. This kind of error is known as the systematic error, because it always pushes our result in the same direction. Systematic errors cannot be discovered with the help of statistical analysis. They are often hard to evaluate or even to detect. In our experiments, we often (but not always) assume that systematic errors have been made much smaller than the required precision.

2. Significant figures

In the system with base 10 (called the decimal system), a number is a composition of ten digits 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. We generally deal with two types of numbers: (i) integers, and (ii) real numbers. An integer is a signed or unsigned whole numbers without a decimal point or fraction. Examples of integers are: 25, 100, −375, −6789, etc. A real number in general has a decimal point. It is the decimal point which helps in distinguishing between a real number and an integer. For example, 47 is an integer and 47. is a real number. The real numbers may be expressed either in the fractional from or in the exponential form (also known as the scientific notation). Some examples of real numbers are: 25.6700, −70.05, 2.83×10^2, 6.02×10^{-3}, -7.56×10^4, -8.004×10^{-9}, etc. Here, the first two numbers are expressed in the fractional from whereas the next four numbers are expressed in the exponential form.

Whether all the digits in a given number are significant or not, let us look into it. A significant figure or significant digit in a number is any one of the digits 1, 2, 3, . . . , 9. The digit 0 is a significant figure except when it is used to fix the decimal point or to fill the places of unknown or discarded digits. For example, in the number 0.00678, the significant figures are 6, 7, and 8. Here, the zeroes are not significant figures. In the number 4507, all the digits including the zero, are significant figures. The ambiguity for deciding about the significant figures in a number can be sorted out by expressing a number in the scientific notation. Then the number of significant figures is indicated by the factor at the left. For example, in the number 356000, the number of significant figures may be ambiguous, whereas the numbers 3.56×10^5, 3.560×10^5, 3.5600×10^5 have three, four and five significant figures, respectively. The number 0.006087 can be expressed as 6.087×10^{-3}. Then the significant figures are 6, 0, 8, and 7. Furthermore, in the numbers 2.487×10^{-3}, 2.4870×10^{-3}, 2.48700×10^{-3}, the number of significant figures are four, five and six, respectively.

Exercise 1: Find the significant figures in the numbers: 0.00405, 4.050, 1.02×10^2, 1.020×10^2, 1.0200×10^2, 3.56×10^{-2}, 3.560×10^{-2}, 3.5600×10^{-2}.

Solution: The significant figures in the numbers are given in the following table.

Number	Significant figures in the number
0.00405	4, 0, 5
4.050	4, 0, 5, 0
1.02×10^2	1, 0, 2
1.020×10^2	1, 0, 2, 0
1.0200×10^2	1, 0, 2, 0, 0
3.56×10^{-2}	3, 5, 6
3.560×10^{-2}	3, 5, 6, 0
3.5600×10^{-2}	3, 5, 6, 0, 0

3. Approximate numbers

A large number of numbers, we deal with, consists of approximate numbers. There are exact numbers also. The numbers such as 6, 4/7, 235, etc. are exact, as there is no approximation or uncertainty associated with them.[1] Though the numbers, such as π, $\sqrt{3}$, e, etc. are exact numbers, they however cannot be expressed exactly in the form of digits. When expressed in the digital form, they must be, for example, written as 3.1416, 1.7321, 2.7183, etc. These numbers are therefore only approximations to the true values[2] and in such cases are called the approximate numbers. An approximate number is therefore a number which is used as an approximation to its exact value and differs only slightly from the exact value for which it stands.

4. Rounding off numbers

In order to understand the process of rounding off the numbers, let us, for example, attempt to divide 23.6 by 15.2 to get

[1]During the execution, the number 4/7 also becomes as an approximate number.
[2]We have $\pi = 3.141592654....$, $\sqrt{3} = 1.732050808....$, $e = 2.718281828....$

$23.6/15.2 = 1.552631579....$

Obviously, this number never terminates. For using such a number in computation, we would like to cut it down to a manageable form, such as 1.55, 1.553, 1.5526, ..., 1.5526316 etc. This process of retaining, as many as desired digits and cutting off less important digits is known as the rounding off. After the rounding off, we obtain an approximate number. While rounding off a number, care is taken that it should cause the least possible error. In order to round off a number to an appropriate value, say, up to nth place, we discard all the digits to the right of the nth place according to the following rules:

(a) When the discarded number is less than half a unit in the nth place, leave the digit at the nth place unchanged.

(b) When the discarded number is more than half a unit in the nth place, add 1 to the digit at the nth place.

(c) When the discarded number is exactly half a unit in the nth place, leave the nth digit unchanged when it is an even number, but increase it by 1 when it is an odd number. Obviously, it will leave an even number in both cases at the nth place.

According to the above rules, some numbers rounded off to five significant digits are given in the following table:

Number	Rounded number
53.67298	53.673
91.89234	91.892
6.48995001	6.4900
46.634479	46.634
19.0465	19.046
47.1475	47.148

When the number is rounded off, it is said to be correct to the nth place. Thus, by rounding off a number, some error is introduced. When the rounding off the numbers is done according to the said rules, in the computation of a mathematical expression, the errors due to rounding off may be largely canceled by one another.

Exercise 2: Round off the numbers 65.602345, −3.289234, 435.0025689, 554.082546, 678.8565 up to one, two and three decimal points.

Solution: The values rounded off up to one, two and three decimal points are given in the following table.

Number	Rounded up to one dec point	Rounded up to two dec points	Rounded up to three dec points
65.602345	65.6	65.60	65.602
−3.289234	−3.3	−3.29	−3.289
435.0025689	435.0	435.00	435.003
554.082546	554.1	554.08	554.083
678.8565	678.9	678.86	678.857

Exercise 3: Round off the numbers 256789.2, −302502, 502460, 674829, 1012.456 up to unit, ten and hundred positions.

Solution: The values rounded off up to unit, ten and hundred positions are given in the following table.

Number	Rounded up to unit position	Rounded up to ten position	Rounded up to hundred position
256789.2	256789	256790	256800
−302502	−302502	−302500	−302500
502460	502460	502460	502500
674829	674829	674830	674800
1012.456	1012	1010	1000

4.1 Further observations

About the rule (c) mentioned above, further care need to be taken as the following.

(i) Let us first look into the old rule which states that when 5 is dropped, the preceding digit should always be increased by 1. It may accumulate the rounding off errors. In order to minimize the rounding off errors, the preceding digit should be increased by 1 in only half cases and should be left unchanged in the other half cases.

The reason for this opinion would become clear from the following simple example of the multiplication of 6.65×3.75 ($= 24.9375 = 24.94$) after rounding off both numbers up to first decimal place. We have

$$6.7 \times 3.8 = 25.46 \tag{2.1}$$

$$6.6 \times 3.8 = 25.08 \tag{2.2}$$

$$6.7 \times 3.7 = 24.79 \tag{2.3}$$

The results in (2.2) and (2.3) are better than that in (2.1).

(ii) When 5 is to be cut off from two or more numbers in a term, the preceding digit should be increased by 1 in half cases and should be left unchanged in the other half cases, regardless whether the preceding digit is even or odd.

The reason for this would become obvious from the following two simple examples of the multiplication. Let us consider multiplication of 6.65×3.85 (= 25.6025 = 25.60) after rounding off both numbers up to the first decimal place. We have

$$6.6 \times 3.8 = 25.08 \tag{2.4}$$

$$6.6 \times 3.9 = 25.74 \tag{2.5}$$

$$6.7 \times 3.8 = 25.46 \tag{2.6}$$

The results in (2.5) and (2.6) are better than that in (2.4). Let us consider another case of multiplication of 3.75×4.35 (= 16.3125 = 16.31 after rounding off both numbers up to first decimal place. We have

$$3.8 \times 4.4 = 16.72 \tag{2.7}$$

$$3.7 \times 4.4 = 16.28 \tag{2.8}$$

$$3.8 \times 4.3 = 16.34 \tag{2.9}$$

The results in (2.8) and (2.9) are better than that in (2.7).

5. Presentation of errors

Error in a number can be expressed in three different forms: (i) Absolute error, (ii) Relative error, and (iii) Percentage error.

5.1 Absolute error

The absolute error in a measured/calculated value of a physical quantity is the difference between the measured/calculated value and its true value. If N denotes the measured/calculated value of a physical quantity, then the absolute error E_a is ΔN. The minimum and maximum possible values of the absolute error are often referred to as the 'limiting errors'. When the measured or calculated value is expressed up to nth place, the maximum possible absolute error is half a unit in the nth place, which can be expressed as

$$E_a = \Delta N = \frac{1}{2}\left(10^{-n}\right)$$

Absolute error has the same dimensions as the physical quantity. Some examples of maximum absolute error are shown in the following table.

N	Max. value of ΔN
5.67	0.005
-0.0023	0.00005
6.75×10^6	0.005×10^6
8.732×10^{-5}	0.0005×10^{-5}
-7.23×10^9	0.005×10^9

Exercise 4: Write down the maximum value of absolute error in the numbers 5.67, 56734, 2600, 3.65×10^{-3}, 4.08×10^3, -6.985×10^{-2}, -5.47×10^4.

Solution: The given number N and maximum value of absolute error ΔN in the number are given in the following table.

N	Max. value of ΔN
5.67	0.005
56734	0.5
2600	0.5
3.65×10^{-3}	0.005×10^{-3}
4.08×10^3	0.005×10^3
-6.985×10^{-2}	0.0005×10^{-2}
-5.47×10^4	0.005×10^4

5.2 Relative error

The relative error for a physical quantity N is expressed as the ratio of the absolute error ΔN and the value N of the quantity. Hence, the relative error E_r is expressed as

$$E_r = \frac{\Delta N}{N}$$

Notice that the relative error is a dimensionless quantity.

5.3 Percentage error

The percentage error E_p for a physical quantity N is expressed as

$$E_p = \frac{\Delta N}{N} \times 100$$

Notice that the percentage error also is a dimensionless quantity.

5.4 Relation between the relative error and absolute error of a physical quantity

The relation between the relative error E_r and absolute error E_a of a physical quantity N is expressed as

$$E_a = E_r \times N$$

6. Index of accuracy

A belief about the accuracy of a result may be that it is represented by the number of decimal points in its reported value. This statement cannot be correct in general. The true index of accuracy of a result is the relative error. It can be understood, for example, with the help of the following exercises.

Exercise 5: In a measurement, diameter of a 2 cm steel shaft is measured to the nearest thousandth of a centimeter. In another measurement, 1 kilometer of railroad track is measured to the nearest cm. Which of the measurements is more accurate.

Solution: Let us denote the two measurements by A and B, respectively. The absolute error in the first measurement is $\Delta A = 0.0005$ cm whereas the absolute error in the second measurement is $\Delta B = 0.5$ cm. According the widespread belief, the measurement A is more accurate as compared to the measurement B. But, it is not the case, as the true index of accuracy is the relative error. The relative errors in the two cases are

$$E_{ra} = \frac{\Delta A}{A} = \frac{0.0005}{2} = \frac{1}{4000}$$

and

$$E_{rb} = \frac{\Delta B}{B} = \frac{0.5}{1 \times 10^5} = \frac{1}{200000}$$

Hence, in the measurement A, there is an error of 1 part in 4000 whereas in the measurement B, there is an error of 1 part in 200000. The latter measurement is clearly more accurate.

Exercise 6: In one measurement, a 5 m rod is measured to the nearest thousandth of a centimeter. In another measurement, 1 kilometer of railroad track is measured to the nearest 1 m. Which of the measurements is more accurate.

Solution: Let us denote the two measurements by A and B, respectively. The absolute error in the first measurement is $\Delta A = 0.0005$ cm whereas the absolute error in the second measurement is $\Delta B = 0.5$ m. In order to decide about the true accuracy, let us calculate the relative errors. The relative errors in the two cases are

$$E_{ra} = \frac{\Delta A}{A} = \frac{0.0005}{5 \times 10^2} = \frac{1}{100000}$$

and

$$E_{rb} = \frac{\Delta B}{B} = \frac{0.5}{1 \times 10^3} = \frac{1}{2000}$$

Hence, in the measurement A, there is an error of 1 part in 100000 whereas in the measurement B, there is an error of 1 part in 2000. The former measurement is clearly more accurate.

7. Error formulas

Let us look into the error formula for some arithmetic operations.

7.1 Addition

Let us have

$$N = u_1 + u_2 + \ldots + u_n$$

Then, we have

$$\Delta N = \Delta u_1 + \Delta u_2 + \ldots + \Delta u_n$$

The maximum probable errors Δu_1, Δu_2, $\ldots$, Δu_n are positive. The absolute error of a sum of approximate numbers is therefore equal to the sum of their absolute errors, .

Exercise 7: Find the sum $S = \sqrt{5} + \sqrt{7} + \sqrt{11} + \sqrt{13}$ up to three decimal places. Calculate the absolute error and relative error in S.

Solution: The values of square roots up to four significant figures are:

$$\sqrt{5} = 2.236 \qquad \sqrt{7} = 2.646 \qquad \sqrt{11} = 3.317 \qquad \sqrt{13} = 3.606$$

Thus, $S = 2.236 + 2.646 + 3.317 + 3.606 = 11.805$. The maximum absolute error in S is

$$\Delta S = 0.0005 + 0.0005 + 0.0005 + 0.0005 = 0.002$$

Hence, there is error at the third decimal place. Hence, the correct value of S is 11.81. The relative error is

$$E_r = \frac{\Delta S}{S} = \frac{0.002}{11.81} = 0.00017$$

Exercise 8: Find the sum of the approximate numbers 478.34, 45.745, 305.4, and 1.0498, each being correct up to its last figure. Calculate the absolute error and relative error in the sum.

Solution: Since the third number is known only to the first decimal place, we should not retain more than two decimal places in any of the other numbers. Thus, we round them off to two decimal places, add them, and give the result to one decimal place, as

$$S = 478.34 + 45.74 + 305.4 + 1.05 = 830.53$$

The absolute error in the sum S is

$$\Delta S = 0.005 + 0.005 + 0.05 + 0.005 = 0.065$$

It shows that there is error at the second decimal place. Thus, the correct value of S is 830.5. The relative error is

$$E_r = \frac{\Delta S}{S} = \frac{0.065}{830.5} = 0.00008$$

Exercise 9: Find the sum of 47890, 892, 534.34, 38400, and 29645, assuming that the number 38400 is known to only three significant figures. Calculate the absolute error and relative error in the sum.

Solution: As one of the numbers, 38400, is known only to the nearest hundreds, we should round off the other numbers to the nearest tens, add them, and give the sum to hundreds, as

$$S = 47890 + 890 + 530 + 38400 + 29640 = 117350$$

The maximum absolute error is

$$\Delta S = 4 \times 5 + 50 = 70$$

Hence, there is an error at the ten position. Hence, the correct value of S is 117400 or 1.174 $\times 10^5$. The relative error is

$$E_r = \frac{\Delta S}{S} = \frac{70}{117400} = 0.0006$$

7.2 Subtraction

Let us have

$$N = u_1 - u_2$$

Then we have

$$\Delta N = \Delta u_1 - \Delta u_2$$

The maximum probable errors Δu_1 and Δu_2 are positive. We must take the sum of the absolute values of errors in order to get the maximum possible error. Thus, we have

$$\Delta N = \Delta u_1 + \Delta u_2$$

Remark: Notice that this expression is the same as would have been for the sum of two numbers u_1 and u_2.

Exercise 10: Subtract 36.346 from 874.2, assuming that each number is approximate and correct only to its last figure.

Solution: Since one figure is up to one decimal place, we round off the other figure up to two decimal places. Now, the value of the difference D is

$$V = 874.2 - 36.35 = 837.85$$

The maximum absolute error is

$$\Delta V = 0.05 + 0.005 = 0.055$$

Hence, there is error at the second decimal place. Hence, the correct value of V is 837.9.

7.3 Multiplication

Let us have

$$N = u_1\ u_2\ \ldots\ u_n$$

On taking logarithm, we have

$$\ln N = \ln u_1 + \ln u_2 + \ldots + \ln u_n$$

On differentiating it, we get

$$\frac{\Delta N}{N} = \frac{\Delta u_1}{u_1} + \frac{\Delta u_2}{u_2} + \ldots + \frac{\Delta u_n}{u_n}$$

The maximum probable errors Δu_1, Δu_2, $\ldots$, Δu_n are positive. Hence, the relative error in the quantity N is

$$E_r = \frac{\Delta u_1}{u_1} + \frac{\Delta u_2}{u_2} + \ldots + \frac{\Delta u_n}{u_n}$$

The relative error of a product of approximate numbers is therefore equal to the arithmetic sum of the maximum possible relative errors of the separate numbers.

7.4 Division

Let us have

$$N = \frac{u_1}{u_2}$$

On taking logarithm, we have

$$\ln N = \ln u_1 - \ln u_2$$

On differentiating it, we get

$$\frac{\Delta N}{N} = \frac{\Delta u_1}{u_1} - \frac{\Delta u_2}{u_2}$$

The maximum probable errors Δu_1 and Δu_2 are positive. We must take them with the positive sign in order to be sure of the maximum error in the function N. Hence, the maximum possible relative error in the quantity N is

$$E_r = \frac{\Delta u_1}{u_1} + \frac{\Delta u_2}{u_2}$$

Exercise 11: Evaluate the product 491.2×563.5 and find out how many figures in the result are trustworthy.

Solution: Since both the numbers, say, u_1 and u_2, are up to one decimal place, we have $\Delta u_1 = 0.05$ and $\Delta u_2 = 0.05$. Hence, the maximum relative error is

$$E_r = \frac{\Delta u_1}{u_1} + \frac{\Delta u_2}{u_2} = \frac{0.05}{491.2} + \frac{0.05}{563.5} = 0.00019$$

The product of the given numbers is $491.2 \times 563.5 = 276791.2$. The absolute error in the product is

$$E_a = 276791.2 \times 0.00019 = 52.59 = 53$$

Thus, the error is in tens. The trustworthy results of the product is $276800 = 2.768 \times 10^5$. Even then there is some uncertainty in the last figure.

Exercise 12: Find out the value of the product 679.2×3.06 and find out how many figures in the result are trustworthy.

Solution: Denoting the terms by u_1 and u_2, respectively, we have $\Delta u_1 = 0.05$ and $\Delta u_2 = 0.005$. Hence, the maximum relative error is

$$E_r = \frac{\Delta u_1}{u_1} + \frac{\Delta u_2}{u_2} = \frac{0.05}{679.2} + \frac{0.005}{3.06} = 0.0017$$

The product of the given numbers is $679.2 \times 3.06 = 2078.352$. The absolute error in the product is

$$E_a = 2078.352 \times 0.0017 = 3.533$$

Thus, the error is in the units. The trustworthy results of the product is $2080 = 2.08 \times 10^3$. Even then there is some uncertainty in the last figure.

7.5 Powers and roots

Here, we have

$$N = u^m$$

where m is a constant. When m is a positive integer, it is the case of the power, and when m is $1/r$, where r is a positive integer, it is the case of the root. Now, the maximum relative error is obtained, by taking logarithm and differentiating, as

$$E_r = \frac{\Delta N}{N} = m\,\frac{\Delta u}{u}$$

Exercise 13: For the quantity N expressed as

$$N = \frac{K a^m b^n c^p}{d^q f^r}$$

where K is some constant, m, n, p, q, and r are powers and roots, and a, b, c, d, and f are measured or calculated values of parameters. Obtain an expression for the maximum relative error in N.

Solution: Relative error for the given expression N can be obtained in the following manner. On taking logarithm, we have

$$\ln N = \ln K + m \ln a + n \ln b + p \ln c - q \ln d - r \ln f$$

On differentiating, we get

$$\frac{\Delta N}{N} = m\,\frac{\Delta a}{a} + n\,\frac{\Delta b}{b} + p\,\frac{\Delta c}{c} - q\,\frac{\Delta d}{d} - r\frac{\Delta f}{f}$$

Now, m, n, p, q, r, Δa, Δb, Δc, $\ldots$, Δe are positive. We must take all the terms with the positive sign in order to get the maximum error in the expression N. Hence, we write

$$E_r = m\,\frac{\Delta a}{a} + n\,\frac{\Delta b}{b} + p\,\frac{\Delta c}{c} + q\,\frac{\Delta d}{d} + r\,\frac{\Delta f}{f}$$

Exercise 14: Calculate the quantity N expressed as

$$N = \frac{36.5^2 \times 47.65 \times 48.35^{1/3}}{45.83^2 \times 83.2^{1/2}}$$

Solution: The calculated value of N is 12.07126. The expression for relative error is

$$\frac{\Delta N}{N} = 2\,\frac{0.05}{36.5} + \frac{0.005}{47.65} + \frac{1}{3}\,\frac{0.005}{48.35} + 2\,\frac{0.005}{45.83} + \frac{1}{2}\,\frac{0.05}{83.2} = 3.398 \times 10^{-3}$$

Hence,

$$\Delta N = (3.398 \times 10^{-3}) \times (12.07126) = 0.041$$

As the error is at the second decimal place, we have $N = 12.1$.

Exercise 15: For the quantity N expressed as

$$N = \frac{K(p+q)^m\ s^p}{(w-x-y)^q (z+u-v)^r}$$

where K is some constant, m, p, q, and r are powers and roots, and p, q, s, w, x, y, z, u and v are measured or calculated values of parameters. Obtain an expression for the maximum relative error in N.

Solution: Let us first express

$$a = p + q \qquad b = w - x - y \qquad c = z + u - v$$

so that

$$N = \frac{K a^m \; s^p}{b^q \; c^r}$$

On taking logarithm, we have

$$\ln N = \ln K + m \ln a + p \ln s - q \ln b - r \ln c$$

On differentiating, we get

$$\frac{\Delta N}{N} = m \frac{\Delta a}{a} + p \frac{\Delta s}{s} - q \frac{\Delta b}{b} - r \frac{\Delta c}{c}$$

Now, m, p, q, r, Δa, Δs, Δb, Δc are positive. We must take all the terms with the positive sign in order to get the maximum error in the expression N. Hence, we write

$$E_r = m \frac{\Delta a}{a} + p \frac{\Delta s}{s} + q \frac{\Delta b}{b} + r \frac{\Delta c}{c}$$

Here, we have

$$\Delta a = \Delta p + \Delta q \qquad \Delta b = \Delta w + \Delta x + \Delta y \qquad \Delta c = \Delta z + \Delta u + \Delta v$$

Exercise 16: Calculate the quantity N expressed as

$$N = \frac{(15.3 + 20.12)^2 \times (73.85 - 33.2)^3}{(22.2 - 2.56)^2 \times (55.2 + 32.56 - 10.2)^{1/2}}$$

Solution: The calculated value of N is 2.1847×10^5. The relative error is

$$\frac{\Delta N}{N} = 2 \frac{0.05 + 0.005}{35.42} + 3 \frac{0.005 + 0.05}{40.65} + 2 \frac{0.05 + 0.005}{19.64}$$
$$+ \frac{1}{2} \frac{0.05 + 0.005 + 0.05}{77.56} = 0.01344$$

Therefore,

$$\Delta N = 2.1847 \times 10^5 \times 0.01344 = 0.029 \times 10^5$$

For the data expressed in 10^5, the error is at the second decimal place, Thus, we have $N = 2.2 \times 10^5$.

8. Problems and questions

1. What are integers, and real numbers. Discuss about the significant figures in a number. Write down the significant figures in the numbers, 34.5070, -62.0500, 0.00283, 16.026×10^{-3}, -7.056×10^{4}, -8.400×10^{-9}.

2. Explain the process of rounding off the numbers. Round off the numbers 53.67298, 91.89234, 2.48995001, 46.634479, 19.0465 up to two decimal places.

3. What are absolute error, relative error and percentage error? For the measured length 25.38 cm, calculate absolute error, relative error and percentage error.

4. How to compare the accuracy of two measurements. In a measurement, height of a 2 m wall is measured to the nearest tenth of a centimeter. In another measurement, 1 kilometer road is measured to the nearest 1 cm. Which of the measurements is more accurate.

5. Evaluate the following expressions to the best possible accuracy:

   (a) $638.242 + 59.2233 + 472.2 + 3.0507 + 0.465$

   (b) $874.2 - 36.345$

   (c) 506.4×25.66

   (d) $(626.4 \times 25.66)/(34.5 \times 90.094)$

   (e) $(654.3 \times 13.25)/(38.2 \times 67.452) + (9.993 \times 2.45)/(47.2 \times 33.033)$

6. Write short notes on the following:

   (i) Random and systematic errors

   (ii) Significant figures in a number

   (iii) Approximate numbers

   (iv) Approximate numbers and significant figures

   (v) Rounding of numbers

   (vi) Absolute, relative and percentage errors

III. Arrangement of Data

Data are considered as raw material for any investigation. In an survey about the age of persons, information about 264 persons was collected. The persons are named here by assigning them a number in sequence. The ages of the persons are given in Table 1. This complete set of data is often referred to as the population. Total number of data in the population is denoted by N. Here N is 264.

This is a data set of moderate size. The data set may be quite big. In the present time of computers, dealing with them has become quite manageable as the only requirement is to enter the data and other work is generally performed with the help of computer softwares developed for the purpose. But, during early days when computers were not available, treatment of large data set was unmanageable and people had to select sample out of the data.

1. Sampling

Sampling of data has been a common practice in the early days when computers were not accessible. Sampling may be required when the data set is quite large. The sample can be taken in an random manner or in a systematic manner. The number of data in a sample is denoted by n. Obviously, a sample is a subset of the population.

1.1 Random sampling

In the random sampling, we select the data out of the population in a random manner. For example, out of 264 data given in Table 1, ten data selected in a random manner out of them are shown in the following table.

P =	12	135	233	143	195	68	104	47	254	77
Age x =	78	22	38	44	46	28	17	54	62	31

Table 1. Ages of persons

P	Age	P	Age	P	Age	P	Age	P	Age	P	Age
1	49	45	33	89	54	133	28	177	67	221	83
2	64	46	15	90	53	134	46	178	68	222	44
3	24	47	54	91	81	135	22	179	27	223	38
4	30	48	49	92	52	136	33	180	44	224	34
5	77	49	61	93	72	137	89	181	46	225	21
6	45	50	67	94	85	138	48	182	54	226	21
7	15	51	32	95	53	139	88	183	56	227	49
8	36	52	62	96	42	140	30	184	51	228	60
9	22	53	57	97	81	141	33	185	64	229	43
10	29	54	19	98	74	142	25	186	62	230	36
11	25	55	38	99	30	143	44	187	41	231	16
12	78	56	31	100	74	144	64	188	39	232	19
13	45	57	86	101	39	145	50	189	42	233	38
14	46	58	23	102	86	146	63	190	22	234	31
15	25	59	61	103	70	147	33	191	54	235	79
16	20	60	43	104	17	148	36	192	78	236	61
17	84	61	71	105	22	149	52	193	19	237	36
18	72	62	48	106	81	150	33	194	58	238	32
19	63	63	34	107	71	151	50	195	46	239	59
20	78	64	84	108	22	152	26	196	40	240	58
21	62	65	47	109	50	153	74	197	38	241	53
22	28	66	65	110	19	154	67	198	75	242	81
23	38	67	26	111	18	155	52	199	49	243	44
24	69	68	28	112	36	156	29	200	35	244	76
25	24	69	37	113	75	157	82	201	16	245	17
26	23	70	87	114	80	158	77	202	73	246	66
27	82	71	47	115	34	159	56	203	57	247	61
28	40	72	41	116	58	160	35	204	37	248	24
29	35	73	46	117	57	161	63	205	71	249	77
30	84	74	65	118	59	162	42	206	33	250	29
31	22	75	19	119	18	163	17	207	23	251	34
32	45	76	56	120	67	164	78	208	89	252	16
33	22	77	31	121	44	165	52	209	61	253	78
34	16	78	27	122	38	166	49	210	51	254	62
35	73	79	84	123	25	167	18	211	29	255	31
36	42	80	53	124	64	168	75	212	86	256	82
37	25	81	34	125	45	169	56	213	64	257	73
38	82	82	21	126	32	170	43	214	54	258	65
39	76	83	87	127	23	171	33	215	44	259	55
40	66	84	77	128	88	172	15	216	26	260	37
41	48	85	59	129	61	173	72	217	83	261	24
42	35	86	46	130	57	174	68	218	79	262	80
43	27	87	38	131	49	175	57	219	65	263	74

1.2 Systematic sampling

Let us select 9 values out of the population in a systematic manner. It can be done in many ways. For example, starting from person No. 7 and every time adding 30 to the person No., nine data selected in a systematic manner from Table 1 are shown in the following table.

P =	7	37	67	97	127	157	187	217	247
Age =	64	25	26	81	23	82	41	83	61

2. Arrangement of data

For the analysis purpose, one would like to arrange the data in a systematic manner. For example, in the present case, the data given in Table 1 can be arranged in the manner given in Table 2. Here, we have arranged the ages in the increasing manner and have given frequencies f for various ages of persons.

Table 2. Frequencies of various ages of persons

Age	f	Age	f	Age	f	Age	f	Age	f	Age	f
15	3	28	3	41	2	54	5	67	4	80	2
16	4	29	4	42	4	55	1	68	2	81	4
17	3	30	3	43	3	56	4	69	1	82	4
18	3	31	4	44	6	57	6	70	1	83	2
19	5	32	3	45	4	58	3	71	3	84	4
20	1	33	7	46	6	59	3	72	3	85	1
21	3	34	5	47	2	60	1	73	3	86	3
22	7	35	4	48	5	61	6	74	4	87	2
23	4	36	5	49	6	62	4	75	3	88	2
24	4	37	4	50	3	63	3	76	2	89	2
25	5	38	7	51	2	64	5	77	4		
26	4	39	3	52	4	65	4	78	5		
27	3	40	2	53	4	66	2	79	2		

The sum of the ages of all 264 persons is 13005. Thus, the average age is $13005/264 = 49.26$ years.

2.1 Grouped data

Another way of arranging data may be to group them in the form of groups, expressed in the form of intervals, and counting the number of data falling in various intervals, as shown in Table 3 for the data given in Table 1. Here, the intervals are $10 - 19, 20 - 29, \ldots, 80 - 89$. The number of data lying in an interval is called the frequency. The size of intervals is decided by the investigator oneself.

Table 3. Frequencies and relative frequencies

Interval	Freq	Relative freq
10 – 19	18	0.0682
20 – 29	38	0.1439
30 – 39	45	0.1705
40 – 49	40	0.1515
50 – 59	35	0.1326
60 – 69	32	0.1212
70 – 79	30	0.1136
80 – 89	26	0.0985
Sum	264	1.000

This table, for example, shows that age of 40 persons lie in the age group from 40 to 49. Sometimes, it is useful to have information about the relative frequencies. The relative frequency is obtained by dividing the frequency by the total number of data. They are given in column 4 of Table 3.

3. Arithmetic mean

Arithmetic mean (often called, mean) of data x_i, $i = 1, 2, \ldots, n$ is expressed as

$$\overline{x} = \sum_{i=1}^{n} x_i \Big/ n$$

In case of group-intervals, x_i represents the mid-value of the corresponding interval.

Exercise 1: Calculate the mean of 15, 20, 22, 34, 47.

Solution: There are five data, therefore $n = 5$. The mean (or arithmetic mean) is

$$\overline{x} = \frac{\sum x_i}{n} = \frac{15 + 20 + 22 + 34 + 47}{5} = \frac{138}{5} = 27.6$$

4. Median

When we arrange data in a sequence (increasing or decreasing), the middle value in the data set is known as the median of data. When there are two values in the middle, average of them is the median of data.

Remark: For a symmetrical distribution of data, the mean and median of the data set are equal.

Exercise 2: Calculate the median of 15, 20, 52, 22, 65, 34, 47.

Solution: Let us first arrange the data in a sequence, say, in the increasing sequence as:

$$55, 20, 22, 34, 47, 82, 65, 78$$

There are seven data and thus the fourth data is in the middle. Hence, the median of data is 34.

Exercise 3: Calculate the median of 55, 15, 30, 52, 82, 65, 44, 47.

Solution: Let us first arrange the data in a sequence, say, in the increasing sequence as:

$$20, 22, 34, 47, 55, 65, 78, 82$$

There are eight data and thus fourth and fifth data, 47 and 55, respectively, are in the middle. Hence, the median of data is

$$\text{Median} = \frac{47 + 55}{2} = 51$$

5. Variance

When the values in a set of data lie close to their mean, the dispersion is less as compared when the values lie far away from their mean. This tendency of data is expressed with the help of a parameter, called the variance, denoted by s^2 and expressed as

$$s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \Big/ (n - 1)$$

Here, n is the number of values in a set. The reason for dividing by $(n - 1)$ rather than n is the theoretical consideration referred to as the degree of freedom. From the practical point of view, dividing the squared differences by $(n - 1)$ rather than n is necessary in order to use sample variance. The population variance, denoted by σ^2, is expressed as

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \Big/ n$$

6. Standard deviation

The variance represents squared units and, therefore, is not an appropriate measure of dispersion when we wish to express this concept in terms of the original units. To obtain the measure of dispersion in original units, we merely take the square root of the variance. The result is known as the standard deviation, denoted by s

$$s = \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \Big/ (n-1)}$$

It is the standard deviation of a sample. The standard deviation of a finite population is

$$\sigma = \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \Big/ n}$$

Note: When the number n is very large, s and σ are almost equal to each other.

Exercise 4: Calculate standard deviation for the following data:

P =	1	2	3	4	5	6	7	8	9	10
Age =	78	22	38	44	46	28	17	54	62	31

Solution: The mean of the sample is

$$\overline{x} = \frac{\sum x}{n} = \frac{420}{10} = 42$$

For calculation of standard deviation, we prepare a table:

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
78	36	1296
22	−20	400
38	−4	16
44	2	4
46	4	16
28	−14	196
17	−25	625
54	12	144
62	20	400
31	−11	121
420		3218

The standard deviation is

$$\sigma = \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \Big/ n} = \sqrt{\frac{3218}{10}} = 17.94$$

Exercise 5: Calculate standard deviation for the following data:

P =	1	2	3	4	5	6	7	8	9
Age =	64	25	26	81	23	82	41	83	61

Solution: The mean of the sample is

$$\overline{x} = \frac{\sum x}{n} = \frac{486}{9} = 54$$

For calculation of standard deviation, we prepare a table:

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
64	10	100
25	−29	841
26	−28	784
81	27	729
23	−31	961
82	28	784
41	−13	169
83	29	841
61	7	49
486		5258

The standard deviation is

$$\sigma = \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \Big/ n} = \sqrt{\frac{5258}{9}} = 24.17$$

6.1 Coefficient of variance

The coefficient of variance of a sample is expressed as

$$\text{C. V.} = \frac{s}{\overline{x}}\, 100\%$$

Exercise 6: For the values 10, 15, 18, 23, 36, 42 calculate, arithmetic mean, standard deviation and the coefficient of variance.

Solution: We arrange the data in the form of the following table.

No.	x	$(x-\overline{x})$	$(x-\overline{x})^2$
1	10	−14	196
2	15	−9	81
3	18	−6	36
4	23	−1	1
5	36	12	144
6	42	18	324
sum =	144		782

The arithmetic mean of data is

$$\overline{x} = \frac{\sum x_i}{n} = \frac{144}{6} = 24$$

The standard deviation is

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{782}{5}} = 11.416$$

The coefficient of variance is

$$\text{C. V.} = \frac{s}{\overline{x}}\ 100\% = \frac{11.416}{24} \times 100 = 47.57\%$$

7. Probability distribution

The probability distribution $p(x)$ for various situations of an event x is expressed as

$$p(x) = \frac{f}{\sum f}$$

where f denotes the frequency of the situation of the event.

Exercise 7: The frequencies of various situations of an event x are given in the following table. Calculate the probability distribution.

Situation x =	1	2	3	4	5	6	7	8
Frequency f =	18	38	45	40	36	31	29	27

Solution: The sum of all frequencies is 264. We compute the probabilities by dividing their respective frequencies by the total 264. The probabilities are given in column 3 of table.

Table 5. Probability distribution

Situation	Freq	$p(x)$
1	18	0.0682
2	38	0.1439
3	45	0.1705
4	40	0.1515
5	36	0.1364
6	31	0.1174
7	29	0.1098
8	27	0.1023
Sum	264	1.0000

7.1 Properties of probability distribution

It is obvious that the value of a probability $p(x)$ is always positive and less than 1. Further, the sum of all probabilities is equal to 1. These properties may be expressed mathematically as:

(i) $0 \leq p(x) \leq 1$ (ii) $\sum p(x) = 1$

7.2 Cumulative probability distribution

The cumulative distribution can be obtained by dividing the cumulative frequency by the sum. For the given data, the cumulative distribution is shown in the following table.

Table 6. Cumulative probability distribution

Situation	Freq	Cum. Freq	$p(x)$
1	18	18	0.0682
2	38	56	0.2121
3	45	101	0.3826
4	40	141	0.5341
5	36	177	0.6705
6	31	208	0.7879
7	29	237	0.8977
8	27	264	1.0000
Sum	264		1.0000

8. Correlation

Suppose, for in investigation, we have two independent parameters x and y, which are varying simultaneously and their values may be expressed in the form (x_i, y_i), where $i = 1, 2, \ldots, n$. Now, we want to investigate if there is any correlation between the two parameters. The correlation (also called the correlation coefficient) r between the two sets x_i and y_i of the data is expressed as

$$r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 \sum Y_i^2}}$$

with $X_i = x_i - \overline{x}$ and $Y_i = y_i - \overline{y}$ are deviations from their respective means $\overline{x}$ and $\overline{y}$. The value of coefficient r varies in between -1 and $+1$. That is, $-1 \leq r \leq 1$. The value $r = +\,1$, shows that there exists a perfect positive correlation between the two parameters, whereas $r = -1$, shows that there exists a perfect negative correlation between the two parameters. When the two parameters are independent of each other, correlation between them is zero. But, its converse is not true, *i.e.*, if the correlation between two parameters is zero, they are not necessarily independent of each other. Zero correlation shows the absence of linear relationship between the two parameters. Other values of r between -1 and 1 are interpreted accordingly. An alternate formula for computing correlation coefficient is

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

Exercise 8: Calculate the correlation coefficient between the two sets of data x_i and y_i given in the following table.

$x_i =$	5	6	3	8	4	9	7
$y_i =$	7	4	6	5	5	10	9

Solution: Here, $n = 7$. For calculation of the correlation coefficient, we prepare the following table.

	x	y	x^2	y^2	xy
	5	7	25	49	35
	6	4	36	16	24
	3	6	9	36	18
	8	5	64	25	40
	4	5	16	25	20
	9	10	81	100	90
	7	9	49	81	63
sum	42	46	280	332	290

The correlation coefficient is

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

$$= \frac{7 \times 290 - 42 \times 46}{\sqrt{7 \times 280 - (42)^2}\sqrt{7 \times 332 - (46)^2}} = 0.48$$

It shows that there is correlation, but not very good.

Exercise 9: Calculate the correlation coefficient between the two sets of data x_i and y_i given in the following table.

$x_i =$	1.2	1.9	2.3	4.5	3.6	5.3	6.7	2.9
$y_i =$	3.5	5.0	5.5	10.2	8.0	11.5	14.5	6.8

Solution: Here, $n = 8$. For calculation of the correlation coefficient, we prepare the following table.

	x	y	x^2	y^2	xy
	1.2	3.5	1.44	12.25	4.20
	1.9	5.0	3.61	25.00	9.50
	2.3	5.5	5.29	30.25	12.65
	4.5	10.2	20.25	104.04	45.90
	3.6	8.0	12.96	64.00	28.80
	5.3	11.5	28.09	132.25	60.95
	6.7	14.5	44.89	210.25	97.15
	2.9	6.8	8.41	46.24	19.72
sum	28.4	65.0	124.94	624.28	278.87

The correlation coefficient is

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

$$= \frac{8 \times 278.87 - 28.4 \times 65.0}{\sqrt{8 \times 124.94 - (28.4)^2}\sqrt{8 \times 624.28 - (65)^2}} = 1.00$$

It shows that there is very good positive correlation.

Exercise 10: Calculate the correlation coefficient between the two sets of data x_i and y_i given in the following table.

$x_i=$	1.2	1.9	3.5	5.2	6.6	8.2	9.4	10.3	4.6
$y_i=$	2.5	1.2	−2.0	−5.5	−8.4	−11.5	−13.5	−15.5	−4.5

Solution: Here, $n = 9$. For calculation of the correlation coefficient, we prepare the following table.

	x	y	x^2	y^2	xy
	1.20	2.50	1.44	6.25	3.00
	1.90	1.20	3.61	1.44	2.28
	3.50	−2.00	12.25	4.00	−7.00
	5.20	−5.50	27.04	30.25	−28.60
	6.60	−8.40	43.56	70.56	−55.44
	8.20	−11.50	67.24	132.25	−94.30
	9.40	−13.50	88.36	182.25	−126.90
	10.30	−15.50	106.09	240.25	−159.65
	4.60	−4.50	21.16	20.25	−20.70
sum	50.90	−57.20	370.75	687.50	−487.31

The correlation coefficient is

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

$$= \frac{9 \times (-487.31)7 - 50.90 \times (-57.20)}{\sqrt{9 \times 370.75 - (50.90)^2}\sqrt{9 \times 687.50 - (-57.20)^2}} = -1.00$$

It shows that there is very good negative correlation.

9. Regression

Regression is a determination of statistical relationship between two or more variables. Here, one variable is taken as dependent variable and others as independent variables. Obviously, the dependent variable depends on each of the independent variables. When we have only one independent variable, it is known as a simple regression. The basic relationship between the independent variable X and the dependent variable Y may be expressed as

$$\hat{Y} = a + bX$$

where the symbol $\hat{Y}$ denotes the estimated value of Y for a given value of X. This equation is known as the regression relation of Y on X. Here, b is positive for direct relationship and negative for inverse relationship. A generally used method to find the best fit is the least square method. To use it efficiently, we first determine

$$\sum x_i^2 = \sum X_i^2 - n\overline{X}^2$$

$$\sum y_i^2 = \sum Y_i^2 - n\overline{Y}^2$$

$$\sum x_i y_i = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

where n is the number of pairs. Here, $\overline{X}$ represents arithmetic mean of X_i and $\overline{Y}$ represents arithmetic mean of Y_i. Then, we have

$$b = \frac{\sum x_i y_i}{\sum x_i^2} \qquad \text{and} \qquad a = \overline{Y} - b\overline{X}$$

These measures define a and b which will give the best possible fit through the original X and Y points and the value of r can then be worked out as

$$r = \frac{b\sqrt{\sum x_i^2}}{\sqrt{\sum y_i^2}}$$

Exercise 11: Using the following data, for the relation $Y = a + bX$, find the value of the constants a and b. Also calculate the correlation coefficient.

$X =$	4	6	9	13	18	24	32	40
$Y =$	9	12	16	23	30	39	51	63

Solution: We prepare the following table.

X	Y	X^2	Y^2	XY
4	9	16	81	36
6	12	36	144	72
9	16	81	256	144
13	23	169	529	299
18	30	324	900	540
24	39	576	1521	936
32	51	1024	2601	1632
40	63	1600	3969	2520
146	243	3826	10001	6179

Thus, the means $\overline{X}$ and $\overline{Y}$ are

$$\overline{X} = \frac{\sum X}{n} = \frac{146}{8} = 18.25 \qquad \text{and} \qquad \overline{Y} = \frac{\sum Y}{n} = \frac{243}{8} = 30.375$$

Now, we have

$$\sum x_i^2 = \sum X_i^2 - n\overline{X}^2 = 3826 - 8 \times 18.25^2 = 1161.5$$

$$\sum y_i^2 = \sum Y_i^2 - n\overline{Y}^2 = 10001 - 8 \times 30.375^2 = 2619.875$$

$$\sum x_i y_i = \sum X_i Y_i - n\overline{XY} = 6179 - 8 \times 18.25 \times 30.375 = 1744.25$$

Thus, we have

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{1744.25}{1161.5} = 1.502$$

and

$$a = \overline{Y} - b\overline{X} = 30.375 - 1.502 \times 18.25 = 2.964$$

The correlation coefficient is

$$r = \frac{b\sqrt{\sum x_i^2}}{\sqrt{\sum y_i^2}} = \frac{1.502\sqrt{1161.5}}{\sqrt{2619.875}} = 1.00$$

9.1 Alternate expression

Alternatively, for fitting a regression equation of the type

$$\hat{Y} = a + bX$$

we can use the relations

$$\sum Y_i = na + b\sum X_i \qquad \sum X_i Y_i = a\sum X_i + b\sum X_i^2$$

On solving these equations, we get

$$a = \frac{\sum Y_i \sum X_i^2 - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

and

$$b = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

Exercise 12: Fit the following set of data through the relation $y = a + bx$ to get the values of a and b.

$x =$	1.2	1.9	3.5	5.2	6.6	8.2
$y =$	5.25	6.85	10.55	14.45	17.65	21.35

Solution: We arrange the data as given in the following table.

	x	y	x^2	xy
	1.20	5.25	1.44	6.30
	1.90	6.85	3.61	13.01
	3.50	10.55	12.25	36.92
	5.20	14.45	27.04	75.14
	6.60	17.65	43.56	116.49
	8.20	21.35	67.24	175.07
sum	26.60	76.10	155.14	422.94

For fitting the data through the expression $y = ax + b$, we have

$$a = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = 2.49$$

and

$$b = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = 2.3$$

10. Multiple regression

When there are two or more than two independent variables, and analysis concerning relationship is known as the multiple correlation and the equation describing such relationship as the multiple regression equation. Here, we consider only two independent variables and one dependent variable. In this situation then multiple regression equation is

$$\hat{Y} = a + b_1 X_1 + b_2 X_2$$

where X_1 and X_2 are two independent variables and Y the dependent variable. We can now express as

$$\sum Y_i = na + b_1 \sum X_{1i} + b_2 \sum X_{2i}$$

$$\sum X_{1i} Y_i = a \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i} X_{2i}$$

$$\sum X_{2i} Y_i = a \sum X_{2i} + b_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^2$$

It may be noted that the number of normal equations depends on the number of independent variables. For k number of independent variables, there are $k + 1$ normal equations. In the multiple regression analysis, the regression coefficients (*viz.*, b_1, b_2, $\ldots$) become less reliable as degree of correlation between the variables (*viz.*, X_1, X_2, $\ldots$) increases.

Such linear simultaneous equations can be solved with the help of matrix manipulation. However, solution of these equations is beyond the scope of this book.

11. Binomial distribution

Binomial distribution is based on the probability of occurrence of an event. Suppose, in any one trial, if p is the probability of success, then $q = 1 - p$ is the probability of failure. For example, when we throw a dice, the probability of getting 5 is $p = 1/6$ and the probability of not getting 5 is $q = 1 - p = 5/6$. When we have n trials, the probability of getting success is expressed with the help of binomial distribution. In n trials, the number m of successes is expressed by binomial distribution

$$B_p(m,n) = {}^nC_m\, p^m q^{n-m}$$

where ${}^nC_m$, known as the binomial coefficient,[1] is expressed as

$${}^nC_m = \frac{n!}{m!(n-m)!} = \frac{n(n-1)\ldots(n-m+1)}{1 \times 2 \times \ldots \times m}$$

Exercise 13: For m successes in n trials, the binomial distribution function is

$$B_p(m,n) = {}^nC_m\, p^m q^{n-m}$$

where p is the probability of success and q the probability of failure in any one trial, such that $p + q = 1$. Show that

$$\sum_{m=0}^{n} B_p(m,n) = 1$$

Solution: For the binomial distribution function

$$B_p(m,n) = {}^nC_m\, p^m q^{n-m}$$

we have

$$\sum_{m=0}^{n} B_p(m,n) = \sum_{m=0}^{n} {}^nC_m\, p^m q^{n-m} = (p+q)^n$$

Since the success probability p is complement to the failure probability q, we have $p + q = 1$ and therefore

$$\sum_{m=0}^{n} B_p(m,n) = 1$$

[1] Binomial coefficients appear in the expansion

$$(p+q)^n = \sum_{m=0}^{n} {}^nC_m\, p^m q^{n-m} = {}^nC_0\, q^n + {}^nC_1\, p\, q^{n-1} + \ldots + {}^nC_n\, p^n$$

Exercise 14: Suppose, we toss 4 coins, calculate the probability of obtaining 0, 1, 2, 3 or 4 heads.

Solution: When we toss one coin, the probability of obtaining head is $p = 1/2$ and the probability of not obtaining head is $q = 1 - p = 1/2$. In $n = 4$ trials, the probability of getting m heads is expressed through the binomial distribution:

$$B_p(m,4) = \ {}^4C_m \left(\frac{1}{2}\right)^m \left(\frac{1}{2}\right)^{4-m} = \ {}^4C_m \left(\frac{1}{2}\right)^4$$

Thus, the probability of getting $m = 0$ head (*i.e.*, no head) is

$$B_p(0,4) = \ {}^4C_0 \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

The probability of getting $m = 1$ head is

$$B_p(1,4) = \ {}^4C_1 \left(\frac{1}{2}\right)^4 = 0.25$$

The probability of getting $m = 2$ heads is

$$B_p(2,4) = \ {}^4C_2 \left(\frac{1}{2}\right)^4 = \frac{3}{8}$$

The probability of getting $m = 3$ heads is

$$B_p(3,4) = \ {}^4C_3 \left(\frac{1}{2}\right)^4 = 0.25$$

The probability of getting $m = 4$ heads is

$$B_p(4,4) = \ {}^4C_0 \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

Exercise 15: Suppose, we throw 3 dices. Calculate the probability of obtaining 0, 1, 2, 3 sixes.

Solution: When we throw one dice, the probability of obtaining six is $p = 1/6$ and the probability of not obtaining six is $q = 1 - p = 5/6$. In $n = 3$ trials, the probability of getting m sixes is expressed through the binomial distribution

$$B_p(m,3) = \ {}^3C_m \left(\frac{1}{6}\right)^m \left(\frac{5}{6}\right)^{3-m}$$

Thus, the probability of getting $m = 0$ six (*i.e.*, no six) is

$$B_p(0,3) = \ {}^3C_0 \left(\frac{5}{6}\right)^3 = 0.579$$

The probability of getting $m = 1$ six is

$$B_p(1,3) = {}^3C_1 \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 = 0.3472$$

The probability of getting $m = 2$ sixes is

$$B_p(2,3) = {}^3C_2 \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) = 0.0694$$

The probability of getting $m = 3$ sixes is

$$B_p(3,3) = {}^3C_3 \left(\frac{1}{6}\right)^3 = 0.0046$$

These distributions are shown in Figure 1.

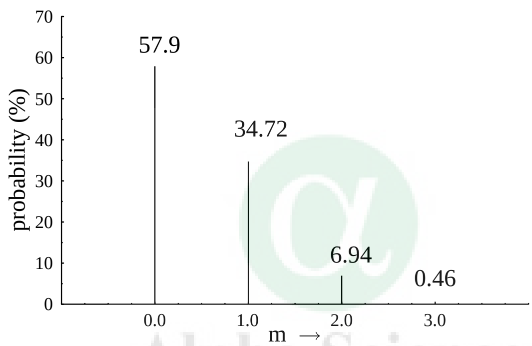


Figure 1: Probability of getting m sixes when throwing 3 dices simultaneously.

Exercise 16: We toss a coin 7 times. How many sequences are for getting 5 times head and 2 times tail.

Solution: If getting head is denoted by p and getting tail by q, then the number of sequences for getting 5 times heads and 2 times tails is the coefficient of p^5q^2 in the expansion of $(p+q)^7$. Thus, it

$${}^7C_5 = \frac{7!}{5!(7-5)!} = \frac{7 \times 6}{2} = 35$$

Exercise 17: In an experiment, the probability of success is 0.1. For $n = 5$ trials, find out the number of sequences and probability $f(x)$ for various numbers x of success.

Solution: For success probability $p = 0.1$, the failure probability $q = 1 - 0.1 = 0.9$. For $n = 5$ trials, a table for the numbers of sequences and probability $f(x)$ for number of successes x are as the following.

No. of successes x	No. of sequences	Probability $f(x)$
0	1	$1 \times (0.1)^5 = 0.00001$
1	5	$5 \times (0.1)^4 \times (0.9) = 0.00045$
2	10	$10 \times (0.1)^3 \times (0.9)^2 = 0.0081$
3	10	$10 \times (0.1)^2 \times (0.9)^3 = 0.0729$
4	5	$5 \times (0.1) \times (0.9)^4 = 0.32805$
5	1	$1 \times (0.9)^5 = 0.59049$
		Total = 1

Exercise 18: For an event, the success is denoted by p and the failure by q. If the event occurs 4 times, write down the binomial probability distribution.

Solution: The binomial distribution is as shown in the following table.

No. of successes x	Probability $f(x)$
0	p^4
1	$4\ p^3q$
2	$6\ p^2q^2$
3	$4\ pq^3$
4	q^4

12. Poisson distribution

Likewise Binomial distribution, Poisson distribution is a discrete distribution. This distribution has been used extensively as a probability model. If x is the number of occurrences of some random event in an interval of time or space (or some volume of matter), the probability that x will occur is expressed as

$$f(x) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} \qquad \text{where} \qquad x = 0, 1, 2, \ldots$$

where λ is known as the parameter of distribution and the average number of occurrences of the random event in the interval (or volume). The symbol e is constant (to four decimals) 2.7183. The following statements describe what is known as the Poisson process:

1. The occurrences of the events are independent. That is, the occurrences of an event in an interval of space or time has no effect on the probability of a second occurrence of the event, in the same, or any other interval.

2. Theoretically, an infinite number of occurrences of the event must be possible in the interval.

3. The probability of the single occurrence of the event in a given interval is proportional to the length of the interval.

4. In any infinitely small portion of the interval, the probability of more than one occurrence of the event is negligible.

An interesting feature of the Poisson distribution is the fact that the mean and variance are equal.

Exercise 19: For the parameter of distribution $\lambda = 2$, evaluate the Poisson distribution and cumulative Poisson distribution for $x = 0$, 1, 2, 3, 4, 5 and 6.

Solution: The Poisson distribution and cumulative Poisson distribution are as shown in the following table.

Value x	Distribution $f(x)$	cumulative distribution
0	0.1353	0.1353
1	0.2707	0.4060
2	0.2707	0.6767
3	0.1804	0.8571
4	0.0902	0.9473
5	0.0361	0.9834
6	0.0120	0.9955

Individual Poisson probabilities for different values of λ are as the following.

x	$\lambda = 0.4$	0.8	1.2	1.5	2	3	5
0	0.6703	0.4493	0.3012	0.2231	0.1353	0.0498	0.0067
1	0.2681	0.3595	0.3614	0.3347	0.2707	0.1494	0.0337
2	0.0536	0.1438	0.2169	0.2510	0.2707	0.2240	0.0842
3	0.0072	0.0383	0.0867	0.1255	0.1804	0.2240	0.1404
4	0.0007	0.0077	0.0260	0.0471	0.0902	0.1680	0.1755
5	0.0001	0.0012	0.0062	0.0141	0.0361	0.1008	0.1755
6	0.0000	0.0002	0.0012	0.0035	0.0120	0.0504	0.1462
7	0.0000	0.0000	0.0002	0.0008	0.0034	0.0216	0.1044
8	0.0000	0.0000	0.0000	0.0001	0.0009	0.0081	0.0653
9	0.0000	0.0000	0.0000	0.0000	0.0002	0.0027	0.0363
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0181

13. Normal distribution

A normal distribution is expressed as

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}}\, \mathrm{e}^{-(x-\mu)^2/2\sigma^2} \tag{3.1}$$

The two parameters of the distribution are the mean μ, and the standard deviation σ, obtained from the given data set. The variable x is continuous and may take on the values in the range from $-\infty$ to ∞. The graph of normal distribution produces the familiar bell-shaped curve shown in figure 2.

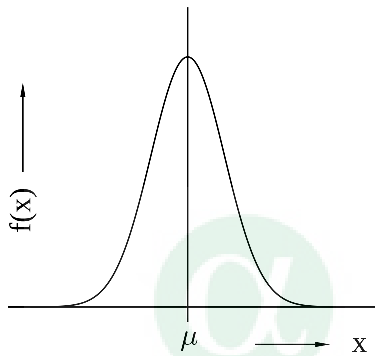


Figure 2: Normal distribution having center at μ.

Some important characteristics of the normal distribution are as the following.

1. It is symmetrical about the mean value μ, *i.e.,* the curve on either side of μ is a mirror image of the other side.

2. The total area under the curve above the x-axis is one square unit. Because of the symmetry, 50 percent of the area is to the right of a perpendicular erected at the mean, and 50 percent area is to the left.

3. When we erect perpendiculars at a distance of 1 standard deviation (*i.e.,* 1 σ) from the mean in both directions, the area enclosed by the perpendiculars, the x-axis and the curve will be 68.26% of the total area. When we extend these lateral boundaries to a distance of 2 σ on either side of the mean, 95.44% of the total area will be enclosed, and on extending them to a distance of 3 σ will cause approximately 99.73% of the total area to be enclosed.

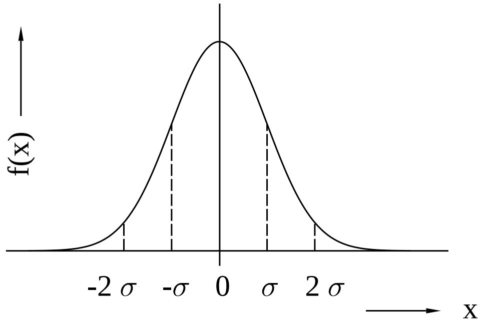


Figure 3: Area between the lines $x = \sigma$ and $x = -\sigma$ is 68.26% and between the lines $x = 2\sigma$ and $x = -2\sigma$ is 95.44%.

4. The normal distribution is specified by the parameters μ and σ. Figure 4 shows three normal distributions with different values of means, but the same value of the standard deviation. Figure 5 shows two normal distributions with different values of standard deviations, but the same value of the mean.

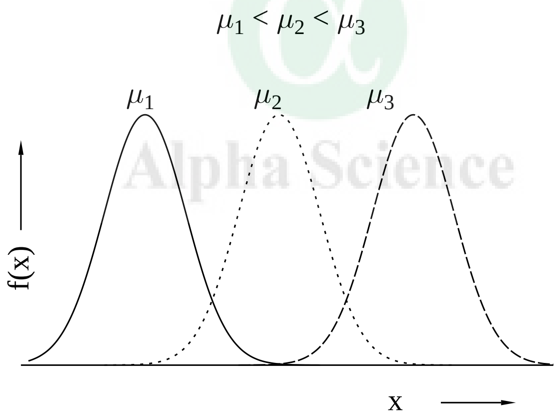


Figure 4: Three normal distributions with different values of means, but the same value of the standard deviation.

13.1 Standard normal distribution

A normal distribution is really a family of distributions in which one member is distinguished from another on the basis of the values of μ and σ. The most important member of this family is the standard normal distribution (also called the unit normal distribution) which has a mean value 0 and standard deviation of 1. It can be obtained from equation 3.1 by creating a random variable $z = (x - \mu)/\sigma$. Thus, the expression

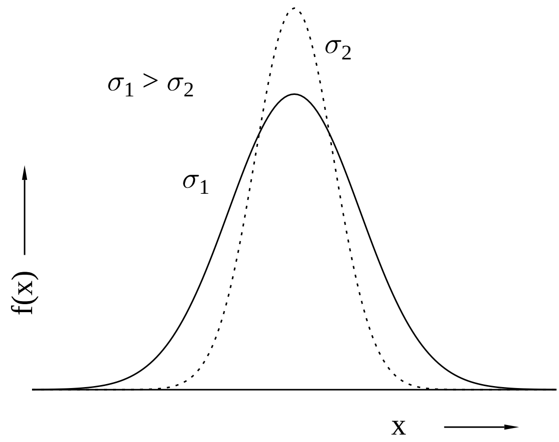


Figure 5: Two normal distributions with different values of standard deviations, but the same value of the mean.

for standard normal distribution is

$$f(z) = \frac{1}{\sqrt{2\pi}}\ \mathrm{e}^{-z^2/2}$$

where z varies from $-\infty$ to ∞. The graph of standard normal distribution is shown in Figure 6. Henceforth, this standard normal distribution would be referred to as the normal distribution.

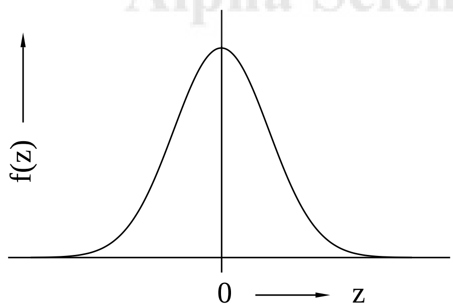


Figure 6: Normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.

To find the probability that the parameter z takes on a value between any two values, say z_0 and z_1, we must find the area bounded by perpendiculars erected at these points, the curve and the horizontal axis. It can also be found from the integral

$$\int_{z_0}^{z_1} \frac{1}{\sqrt{2\pi}}\ \mathrm{e}^{-z^2/2}\ \mathrm{d}z$$

Evaluation of integral

We would like to know the value of the integral

$$\int_{-\infty}^{z_0} \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-z^2/2} \, \mathrm{d}z$$

where z_0 is either a positive or negative real number. It can be evaluated graphically as well as numerically. For the numerical evaluation, for positive value of z_0, we can write as

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} \mathrm{e}^{-z^2/2} \, \mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \mathrm{e}^{-z^2/2} \, \mathrm{d}z + \frac{1}{\sqrt{2\pi}} \int_{0}^{z_0} \mathrm{e}^{-z^2/2} \, \mathrm{d}z$$

$$= 0.5 + \frac{1}{\sqrt{2\pi}} \int_{0}^{z_0} \mathrm{e}^{-z^2/2} \, \mathrm{d}z$$

For negative value of z_0, we can write as

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z_0} \mathrm{e}^{-z^2/2} \, \mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \mathrm{e}^{-z^2/2} \, \mathrm{d}z + \frac{1}{\sqrt{2\pi}} \int_{0}^{-z_0} \mathrm{e}^{-z^2/2} \, \mathrm{d}z$$

$$= 0.5 - \frac{1}{\sqrt{2\pi}} \int_{0}^{z_0} \mathrm{e}^{-z^2/2} \, \mathrm{d}z$$

Calculated values of integral are given in Table 1 in Appendices.

14. *t*-distribution

For a sample having n data, sample standard deviation is expressed as

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \Big/ (n-1)}$$

For population, we had normal distribution and there we defined confidence intervals. For samples having small number of data, it becomes mandatory to find an alternative procedure for constructing confidence intervals. An alternative, known as the Student's t-distribution or t distribution. Percentiles of t-distribution for various confidence intervals and degree-of-freedom are given in Table 2 in Appendices.

15. Problems and questions

1. What is difference between random sampling and systematic sampling.

2. Taking some example, explain about random sampling and systematic sampling.

3. How the population mean and the sample mean are expressed. Calculate with some example.

4. How the population variance and the sample variance are expressed.

5. Calculate standard deviation for the following data:

   62 27 29 82 24 54 86 42 79 55

6. The frequencies of various situations of an event x are given in the following table. Calculate the probability distribution and cumulative probability distribution.

Situation $x =$	1	2	3	4	5	6	7	8
Frequency $f =$	18	38	45	40	36	31	29	27

7. Calculate the correlation coefficient between the two sets of data x_i and y_i given in the following table.

$x_i =$	1.3	1.8	3.5	5.2	5.6	8.2	9.7	9.8	4.8
$y_i =$	2.5	1.2	-2.0	-5.6	-7.5	-8.4	-8.9	-7.6	-3.5

8. Calculate the correlation coefficient between the two sets of data x_i and y_i given in the following table.

$x_i =$	1.2	1.9	2.3	4.5	3.6	5.3	6.7	2.9
$y_i =$	3.5	5.0	5.5	10.2	8.0	11.5	14.5	6.8

9. For an event, the success is denoted by p and the failure by q. If the event occurs 5 times, write down the binomial distribution.

10. Write short notes on the following:

    (i) Arithmetic mean

    (ii) Median

    (iii) Variance

    (iv) Standard deviation

(v) Coefficient of variance

(vi) Probability distribution of discrete values

(vii) Cumulative probability distribution

(viii) Coefficient of variance

(ix) Random and systematic samplings

(x) Population and sample means

(xi) Population and sample variances

(xii) Regression

(xiii) Correlation

(xiv) Multiple correlation and regression

(xv) Binomial distribution

(xvi) Poisson distribution

(xvii) Normal distribution

(xviii) t-distribution

IV. Interpolation

Suppose for an independent variable y and a dependent variable x, we are given a table of the values $(x_i,\, y_i)$, where $i = 1,\, 2,\, \ldots,\, n$, such that $x_1 < x_2 < x_3 < \ldots < x_n$. Interpolation is an art to find out a value in between the lines of the table. For example, suppose we want to find out the value of y corresponding to a given value of x, where $x_j < x < x_{j+1}$, with j as one of the 1, 2, ..., and $n-1$. A procedure to find out the value of y corresponding to the given value of x is known as the interpolation. (When the value of x is outside the range of the given x_i, *i.e.,* either $x < x_1$ or $x > x_n$, the process of finding out the value of y corresponding to x is known as the extrapolation. Results of extrapolation may not be reliable.) In this chapter, we shall discuss about some methods for interpolation. Interpolation of values is a common requirement while dealing with data.

1. Graphical interpolation

A simple method for interpolation, known as graphical method, had been in practice in old days, in particular. It can be understood in the following manner. Define a graph-paper as the (x, y)-plane. Put the given points $(x_i,\, y_i)$, $i = 1,\, 2,\, \ldots,\, n$ on the graph-paper and join these points by a curve smoothly. Now, for finding out a value y' of y corresponding to $x = x'$, draw a line $x = x'$ perpendicular to the x-axis. Intersection of this line and the curve plotted through the given points provides the value y' (which corresponds to x').

Accuracy of this method is equal to that of the graph-paper, which is not very good in comparison to that of the results obtained with the use of a computer. For good accuracy of the results, we go for numerical methods discussed in the following sections.

Exercise 1: Given values of x and the corresponding values of y are

$x =$	0.5	1.2	2.1	2.9	3.6	4.5	5.7
$y =$	3.1	5.0	9.1	14.4	20.3	29.6	45.0

Find out the value of y corresponding to $x = 3.2$ by using the graphical method.

Solution: Figure 1 shows variation of y versus x where the curve is plotted through the given points (x_i, y_i). The line $x = 3.2$ (perpendicular to the x-axis) intersects the curve at $y = 17.1$. Thus, the required value of y is 17.1 corresponding to $x = 3.2$. (Here, the accuracy of the graph-paper along the y-axis is 0.1).

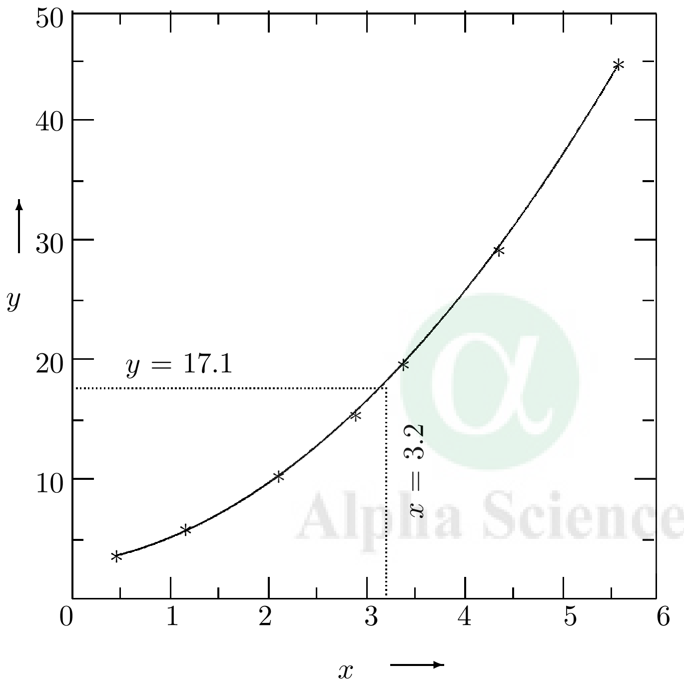


Figure 1: Variation of y versus x.

2. Difference schemes

When the interval between two successive values of x in the data set is constant, interpolation can be done with the help of Gauss forward formula or with the help of Gauss backward formula. For the use of these formulas, we have to prepare forward difference table, preferably.

Suppose we have the values $(x_i,\ y_i)$ $i = 1,\ 2, \ldots,\ n$. Three types of difference schemes are in practice in which the first order differences are defined as:

(a) Forward difference scheme: $\Delta y_i = y_{i+1} - y_i$

(b) Backward difference scheme: $\nabla y_i = y_i - y_{i-1}$

(c) Central difference scheme: $\delta y_i = y_{i+1/2} - y_{i-1/2}$

The second order differences are:

(a) Forward difference scheme: $\Delta^2 y_i = \Delta y_{i+1} - \Delta y_i$

(b) Backward difference scheme: $\nabla^2 y_i = \nabla y_i - \nabla y_{i-1}$

(c) Central difference scheme: $\delta^2 y_i = \delta y_{i+1/2} - \delta y_{i-1/2}$

Higher order differences can be found in the similar manner. For these difference-schemes, the difference tables are given in Tables 1, 2, and 3, respectively.

Table 1. Forward difference table

x	y	Δ	Δ^2	Δ^3	Δ^4	Δ^5	Δ^6	Δ^7	Δ^8
x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$	$\Delta^4 y_0$	$\Delta^5 y_0$	$\Delta^6 y_0$	$\Delta^7 y_0$	$\Delta^8 y_0$
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	$\Delta^4 y_1$	$\Delta^5 y_1$	$\Delta^6 y_1$	$\Delta^7 y_1$	
x_2	y_2	Δy_2	$\Delta^2 y_2$	$\Delta^3 y_2$	$\Delta^4 y_2$	$\Delta^5 y_2$	$\Delta^6 y_2$		
x_3	y_3	Δy_3	$\Delta^2 y_3$	$\Delta^3 y_3$	$\Delta^5 y_3$	$\Delta^6 y_3$			
x_4	y_4	Δy_4	$\Delta^2 y_4$	$\Delta^3 y_4$	$\Delta^4 y_4$				
x_5	y_5	Δy_5	$\Delta^2 y_5$	$\Delta^3 y_5$					
x_6	y_6	Δy_6	$\Delta^2 y_6$						

Table 2. Backward difference table

x	y	∇	∇^2	∇^3	∇^4	∇^5	∇^6	∇^7	∇^8
x_0	y_0								
x_1	y_1	∇y_1							
x_2	y_2	∇y_2	$\nabla^2 y_2$						
x_3	y_3	∇y_3	$\nabla^2 y_3$	$\nabla^3 y_3$					
x_4	y_4	∇y_4	$\nabla^2 y_4$	$\nabla^3 y_4$	$\nabla^4 y_4$				
x_5	y_5	∇y_5	$\nabla^2 y_5$	$\nabla^3 y_5$	$\nabla^4 y_5$	$\nabla^5 y_5$			
x_6	y_6	∇y_6	$\nabla^2 y_6$	$\nabla^3 y_6$	$\nabla^4 y_6$	$\nabla^5 y_6$	$\nabla^6 y_6$		
x_7	y_7	∇y_7	$\nabla^2 y_7$	$\nabla^3 y_7$	$\nabla^4 y_7$	$\nabla^5 y_7$	$\nabla^6 y_7$	$\nabla^7 y_7$	
x_8	y_8	∇y_8	$\nabla^2 y_8$	$\nabla^3 y_8$	$\nabla^4 y_8$	$\nabla^5 y_8$	$\nabla^6 y_8$	$\nabla^7 y_8$	$\nabla^8 y_8$

Table 3. Central difference table

x	y	δ	δ^2	δ^3	δ^4	δ^5	δ^6	δ^7	δ^8
x_0	y_0								
		$\delta y_{1/2}$							
x_1	y_1		$\delta^2 y_1$						
		$\delta y_{3/2}$		$\delta^2 y_{3/2}$					
x_2	y_2		$\delta^2 y_2$		$\delta^4 y_2$				
		$\delta y_{5/2}$		$\delta^3 y_{5/2}$		$\delta^5 y_{5/2}$			
x_3	y_3		$\delta^2 y_3$		$\delta^4 y_3$		$\delta^6 y_3$		
		$\delta y_{7/2}$		$\delta^3 y_{7/2}$		$\delta^5 y_{7/2}$		$\delta^7 y_{7/2}$	
x_4	y_4		$\delta^2 y_4$		$\delta^4 y_4$		$\delta^6 y_4$		$\delta^7 y_4$

Exercise 2: Prepare the forward difference table for the following data.

$x =$	1	2	3	4	5	6	7	8
$y =$	15	10	12	9	14	6	13	7

Solution: The forward difference table for the given data is as the following.

Forward difference table

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$	$\Delta^6 y$	$\Delta^7 y$
1	15	−5	7	−12	25	59	24	156
2	10	2	−5	13	−34	83	−132	
3	12	−3	8	−21	49	−49		
4	9	5	−13	28	0			
5	14	−8	15	28				
6	6	7	−13					
7	13	−6						
8	7							

Exercise 3: Prepare the backward difference table for the following data.

$x =$	11	12	13	14	15	16	17	18
$y =$	21	19	38	16	10	36	43	57

Solution: The backward difference table for the given data is as the following.

Backward difference table

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$	$\nabla^5 y$	$\nabla^6 y$	$\nabla^7 y$
11	21							
12	19	−2						
13	38	19	21					
14	16	−22	−41	−62				
15	10	−6	18	59	−121			
16	36	26	32	14	−45	76		
17	43	7	−19	−51	−65	−20	−96	
18	57	14	7	26	77	142	162	258

3. Gauss formulas for interpolation

The basic requirement for applying Gauss formulas for interpolation is that the interval between successive values of x must be constant, *i.e.*, $x_{i+1}-x_i = x_i - x_{i-1}$. For convenience, we generally consider the forward difference table. Let us consider the following difference table in which the y_0 corresponding to $x = x_0$ is taken some where in the central part of the data set.

Forward difference table

x	y	Δ	Δ^2	Δ^3	Δ^4	Δ^5	Δ^6
x_{-3}	y_{-3}	Δy_{-3}	$\Delta^2 y_{-3}$	$\Delta^3 y_{-3}$	$\Delta^4 y_{-3}$	$\Delta^5 y_{-3}$	$\Delta^6 y_{-3}$
x_{-2}	y_{-2}	Δy_{-2}	$\Delta^2 y_{-2}$	$\Delta^3 y_{-2}$	$\Delta^4 y_{-2}$	$\Delta^5 y_{-2}$	
x_{-1}	y_{-1}	Δy_{-1}	$\Delta^2 y_{-1}$	$\Delta^3 y_{-1}$	$\Delta^4 y_{-1}$		
x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$			
x_1	y_1	Δy_1	$\Delta^2 y_1$				
x_2	y_2	Δy_2					
x_3	y_3						

3.1 Gauss forward formula

For finding out y_p corresponding to x_p, where $x_p = x_0 + p(x_1 - x_0)$, the Gauss' forward formula is

$$y_p = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_{-1} + \frac{p(p^2-1)}{3!}\Delta^3 y_{-1} + \frac{p(p^2-1)(p-2)}{4!}\Delta^4 y_{-2} + \dots$$

Here, x_0 is chosen so that $|p| < 1$. The p is obtained as

$$p = \frac{x_p - x_0}{x_1 - x_0}$$

Exercise 4: Using the following data, obtain the value of y corresponding to $x = 9$ with the help of Gauss' forward interpolation formula.

$x =$	2	5	8	11	14	17
$y =$	6.7	14.2	21.6	29.2	36.9	44.9

Solution: The forward difference table for the given data is as the following.

Forward difference table

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
2	6.7	7.5	−0.1	0.3	−0.4
5	14.2	7.4	0.2	−0.1	0.3
8	21.6	7.6	0.1	0.2	
11	29.2	7.7	0.3		
14	36.9	8.0			
17	44.9				

For $x = 9$, we have $x_0 = 8$ and therefore, we have $p = (9-8)/(11-8) = 1/3$. Form the table

$$y_0 = 21.6, \quad \Delta y_0 = 7.6, \quad \Delta^2 y_{-1} = 0.2, \quad \Delta^3 y_{-1} = -0.1, \quad \Delta^4 y_{-2} = -0.4$$

Using these values, we get

$$y_p = 21.6 + \frac{1}{3}(7.6) - \frac{1}{9}(0.2) - \frac{4}{81}(-0.1) + \frac{5}{243}(-0.4) + \ldots = 24.11$$

3.2 Gauss backward formula

For finding out y_p corresponding to x_p, where $x_p = x_0 + p(x_1 - x_0)$, the Gauss' backward formula is

$$y_p = y_0 + p\Delta y_{-1} + \frac{p(p+1)}{2!}\Delta^2 y_{-1} + \frac{p(p^2-1)}{3!}\Delta^3 y_{-2} + \frac{(p+2)p(p^2-1)}{4!}\Delta^4 y_{-2} + \ldots$$

Here, x_0 is chosen so that $|p| < 1$. The p is obtained as

$$p = \frac{x_p - x_0}{x_1 - x_0}$$

Exercise 5: Using the following data, obtain the value of y corresponding to $x = 15.5$ with the help of Gauss' backward interpolation formula.

$x =$	7	9	11	13	15	17	19
$y =$	10	13	15	18	20	21	25

Solution: The forward difference table for the given data is as the following.

Forward difference table

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
7	10	3	−1	2	−4
9	13	2	1	−2	2
11	15	3	−1	0	4
13	18	2	−1	4	
15	20	1	3		
17	21	4			
19	25				

For $x = 15.5$, we have $x_0 = 15$ and therefore, we have $p = (15.5 - 15)/(17 - 15) = 0.25$. Form the table

$$y_0 = 20, \quad \Delta y_{-1} = 2, \quad \Delta^2 y_{-1} = -1, \quad \Delta^3 y_{-1} = 4, \quad \Delta^4 y_{-2} = 4$$

Using these values, we get

$$y_p = 20 + 0.25 \times 2 + \frac{5}{32}(-1) - \frac{5}{128}(4) - \frac{45}{2048}(4) = 20.01$$

3.3 Limitation of Gauss formulas for interpolation

Gauss formulas for interpolation can only be applied when the separation between the successive values of independent variable x is constant.

4. Linear least square fitting

Suppose, we are given the values (x_i, y_i) where $i = 1, 2, \ldots, n$ and we wish to fit the data through the straight line $y = ax + b$, the values of the constants a and b can be obtained with the help of the following expressions:[1]

$$a = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \qquad \text{and} \qquad b = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

There is no condition of constant separation between successive values of independent variable x to be constant.

4.1 Interpolation through linear least square fitting

After fitting the values (x_i, y_i) through the relation $y = ax+b$, the value y_p corresponding to x_p is obtained as

$$y_p = ax_p + b$$

Exercise 6: Fit the following set of data through the relation $y = ax+b$, and find the value of y corresponding to $x = 2.4$.

[1]Though the derivation of these expression may not be a part of the curriculum for the course on research methodology, but for those who are interested in the derivation, it can be obtained in the following manner. For the linear relation between x and y, we would like to fit the data through the expression

$$f(x) = ax + b$$

with the requirement that $f(x_i)$ should be equal to y_i. The difference between y_i and $f(x_i)$ is

$$d_i = y_i - f(x_i) = y_i - ax_i - b$$

In order to express the largest possible error, we define a parameter S as

$$S = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

Now, we shall adjust a and b so that S is as minimum as possible. That is, we should have

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) = 0 \qquad \text{and} \qquad \frac{\partial S}{\partial b} = \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) = 0$$

These requirements give

$$a\sum x_i^2 + b\sum x_i = \sum x_i y_i \qquad \text{and} \qquad a\sum x_i + nb = \sum y_i$$

On solving these two linear equations for a and b, we get

$$a = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \qquad \text{and} \qquad b = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

$x =$	1	2	3	4
$y =$	1.1	4.2	7.3	10.4

Solution: We arrange the data as given in the following table.

	x	y	x^2	xy
	1	1.1	1	1.1
	2	4.2	4	8.4
	3	7.3	9	21.9
	4	10.4	16	41.6
sum	10	23	30	73

For fitting the data through the expression $y = ax + b$, we have

$$a = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{4 \times 73 - 10 \times 23}{4 \times 30 - (10)^2} = 3.1$$

and

$$b = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{23 \times 30 - 10 \times 73}{4 \times 30 - (10)^2} = -2$$

Thus, we have $y = 3.1x - 2$. For $x = 2.4$, we have $y = 3.1 \times 2.4 - 2 = 5.44$

Exercise 7: Fit the following set of data through the relation $y = ax + b$, and find the value of y corresponding to $x = 3.5$.

$x =$	1	2	3	4	5	6
$y =$	1.0	5.1	9.5	13.5	17.5	22

Solution: We arrange the data as given in the following table.

	x	y	x^2	xy
	1	1	1	1
	2	5.1	4	10.2
	3	9.5	9	28.5
	4	13.5	16	54
	5	17.5	25	87.5
	6	22	36	132
sum	21	68.6	91	313.2

For fitting the data through the expression $y = ax + b$, we have

$$a = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{6 \times 313.2 - 21 \times 68.6}{6 \times 91 - (21)^2} = 4.18$$

and

$$b = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{68.6 \times 91 - 21 \times 313.2}{6 \times 91 - (21)^2} = -3.19$$

Thus, we have $y = 4.18x - 3.19$. For $x = 3.5$, we have $y = 4.18 \times 3.5 - 3.19 = 11.44$

Exercise 8: Fit the following set of data through the relation $y = ax+b$, and find the value of y corresponding to $x = 2.4$.

$x =$	1.2	1.9	3.5	5.2	6.6	8.2
$y =$	5.25	6.85	10.55	14.45	17.65	21.35

Solution: We arrange the data as given in the following table.

	x	y	x^2	xy
	1.20	5.25	1.44	x 6.30
	1.90	6.85	3.61	13.01
	3.50	10.55	12.25	36.92
	5.20	14.45	27.04	75.14
	6.60	17.65	43.56	116.49
	8.20	21.35	67.24	175.07
sum	26.60	76.10	155.14	422.94

For fitting the data through the expression $y = ax + b$, we have

$$a = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{6 \times 422.94 - 26.60 \times 76.10}{6 \times 155.14 - (26.60)^2} = 2.3$$

and

$$b = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{76.10 \times 155.14 - 26.60 \times 422.94}{6 \times 155.14 - (26.60)^2} = 2.5$$

Thus, we have $y = 2.3x + 2.5$. For $x = 2.4$, we have $y = 2.3 \times 2.4 + 2.5 = 8.02$.

5. Curvilinear least square fitting

Quite often it is not possible to fit the data through the linear relation, $y = ax + b$. Then we try to fit them through some non-linear or curvilinear expressions of the form

$$y = b\,\mathrm{e}^{ax}, \qquad y = b\,x^a, \qquad \text{or} \qquad y = a_0 + a_1 x + a_2 x^2 + \ldots$$

where a's and b are constants. The last expression is out of the scope of this book. We shall discuss how to get the least square fitting for the first two expressions.

(i) For the expression

$$y = b\ e^{ax}$$

we take the logarithm to get

$$\ln\ y = \ln\ b + ax$$

It can be expressed as

$$y' = ax + b'$$

where $y' = \ln\ y$ and $b' = \ln\ b$. This is a linear relation as discussed in the preceding section and we have

$$a = \frac{n\Sigma x_i y_i' - \Sigma x_i \Sigma y_i'}{n\Sigma x_i^2 - (\Sigma x_i)^2} \qquad \text{and} \qquad b' = \frac{\Sigma y_i' \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i'}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

Now, we have $b = \exp(b')$. In this manner we obtain values of a and b.

Exercise 9: Fit the following set of data through the relation $y = b\ \exp(ax)$, and find the value of y corresponding to $x = 2.4$.

$x =$	1.2	1.9	3.5	5.2	6.6	8.2
$y =$	2.13	1.98	1.69	1.43	1.24	1.06

Solution: We arrange the data as given in the following table.

	x	y	y'	x^2	xy'
	1.20	2.13	0.76	1.44	0.91
	1.90	1.98	0.68	3.61	1.30
	3.50	1.69	0.52	12.25	1.84
	5.20	1.43	0.36	27.04	1.86
	6.60	1.24	0.22	43.56	1.42
	8.20	1.06	0.06	67.24	0.48
sum	26.60		2.60	155.14	7.80

Here, y' is $\ln\ y$. For fitting the data through the expression $y' = ax + b'$, we have

$$a = \frac{n\Sigma x_i y_i' - \Sigma x_i \Sigma y_i'}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{6 \times 7.80 - 26.60 \times 2.60}{6 \times 155.14 - (26.60)^2} = -0.10$$

and

$$b' = \frac{\Sigma y_i' \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i'}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{2.60 \times 155.14 - 26.60 \times 7.80}{6 \times 155.14 - (26.60)^2} = 0.874$$

Thus, $b = \exp(0.874) = 2.40$, we have $y = 2.40\ \exp(-0.1x)$. For $x = 2.4$, we have $y = 2.40\ \exp(-0.1 \times 2.4) = 1.89$.

(ii) For the expression

$$y = b\, x^a$$

we take the logarithm to get

$$\ln y = \ln b + a \ln x$$

It can be expressed as

$$y' = a\, x' + b'$$

where $y' = \ln y$, $x' = \ln x$ and $b' = \ln b$. This is a linear relation as discussed in the preceding section and we have

$$a = \frac{n\Sigma x'_i y'_i - \Sigma x'_i \Sigma y'_i}{n\Sigma x'^2_i - (\Sigma x'_i)^2} \qquad \text{and} \qquad b' = \frac{\Sigma y'_i \Sigma x'^2_i - \Sigma x'_i \Sigma x'_i y'_i}{n\Sigma x'^2_i - (\Sigma x'_i)^2}$$

Now, we have $b = \exp(b')$.

Exercise 10: Fit the following set of data through the relation $y = b\, x^a$, and find the value of y corresponding to $x = 2.4$.

$x =$	1.2	1.9	3.5	5.2	6.6	8.2
$y =$	2.92	5.30	11.72	19.61	26.74	35.46

Solution: We arrange the data as given in the following table.

x	y	x'	y'	x'^2	$x'y'$
1.20	2.92	0.18	1.07	0.03	0.20
1.90	5.30	0.64	1.67	0.41	1.07
3.50	11.72	1.25	2.46	1.57	3.08
5.20	19.61	1.65	2.98	2.72	4.91
6.60	26.74	1.89	3.29	3.56	6.20
8.20	35.46	2.10	3.57	4.43	7.51
	sum	7.72	15.03	12.72	22.97

Here, y' is $\ln y$ and x' is $\ln x$. For fitting the data through the expression $y' = ax' + b'$, we have

$$a = \frac{n\Sigma x'_i y'_i - \Sigma x'_i \Sigma y'_i}{n\Sigma x'^2_i - (\Sigma x'_i)^2} = \frac{6 \times 22.97 - 7.72 \times 15.03}{6 \times 12.72 - (7.72)^2} = 1.30$$

and

$$b' = \frac{\Sigma y'_i \Sigma x'^2_i - \Sigma x'_i \Sigma x'_i y'_i}{n\Sigma x'^2_i - (\Sigma x'_i)^2} = \frac{15.03 \times 12.72 - 7.72 \times 22.97}{6 \times 12.72 - (7.72)^2} = 0.834$$

Thus, $b = \exp(0.834) = 2.30$, we have $y = 2.30\, x^{1.3}$. For $x = 2.4$, we have $y = 2.30\, (2.4)^{1.3} = 7.18$.

6. Lagrange's interpolation

Suppose we have values (x_i, y_i) $i = 1, 2, \ldots, n$. Lagrange's interpolation formula is

$$y = \sum_{p=1}^{n} \frac{(x - x_1)(x - x_2)\ldots(x - x_n)}{(x_p - x_1)(x_p - x_2)\ldots(x_p - x_n)}\, y_p$$

where the numerator does not contain the term $(x - x_p)$ and the denominator does not contain the term $(x_p - x_p)$. Derivation of this formula is beyond the scope of this book. Here, also there is no restriction for constant separation between the successive values of independent variable x.

Exercise 11: With the help of the Lagrange's interpolation formula obtain the value of y corresponding to $x = 1.5$ from the following data.

$x =$	1	2	3	4
$y =$	1.1	4.3	7.5	10.7

Solution: The Lagrange's interpolation formula is

$$y = \frac{(x - x_2)(x - x_3)(x - x_4)}{(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)}\, y_1 + \frac{(x - x_1)(x - x_3)(x - x_4)}{(x_2 - x_1)(x_2 - x_3)(x_2 - x_4)}\, y_2$$

$$+ \frac{(x - x_1)(x - x_2)(x - x_4)}{(x_3 - x_1)(x_3 - x_2)(x_3 - x_4)}\, y_3 + \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_4 - x_1)(x_4 - x_2)(x_4 - x_3)}\, y_4$$

On substituting the values, we get

$$y = \frac{(1.5 - 2)(1.5 - 3)(1.5 - 4)}{(1 - 2)(1 - 3)(1 - 4)}\, 1.1 + \frac{(1.5 - 1)(1.5 - 3)(1.5 - 4)}{(2 - 1)(2 - 3)(2 - 4)}\, 4.3$$

$$+ \frac{(1.5 - 1)(1.5 - 2)(1.5 - 4)}{(3 - 1)(3 - 2)(3 - 4)}\, 7.5 + \frac{(1.5 - 1)(1.5 - 2)(1.5 - 3)}{(4 - 1)(4 - 2)(4 - 3)}\, 10.7 = 2.7$$

7. Problems and questions

1. What is interpolation? Describe the extrapolation.

2. Explain the graphical method for interpolation.

3. Prepare the forward difference table for the following data.

$x =$	1	2	3	4	5	6	7	8
$y =$	16	13	12	8	14	7	14	9

4. Prepare the backward difference table for the following data.

$x =$	1	2	3	4	5	6	7	8
$y =$	18	13	12	8	19	8	15	9

5. Using the following data, obtain the value of y corresponding to $x = 6$ with the help of Gauss' forward interpolation formula.

$x =$	2	5	8	11	14	17
$y =$	7.8	15.5	22.4	30.3	37.9	46.0

6. Using the following data, obtain the value of y corresponding to $x = 18$ with the help of Gauss' backward interpolation formula.

$x =$	7	9	11	13	15	17	19
$y =$	12	15	16	21	25	24	28

7. Discuss about the interpolation through the linear least square fitting.

8. Fit the following set of data through the relation $y = ax + b$, and find the value of y corresponding to $x = 3.2$.

$x =$	1	2	3	4
$y =$	2.1	5.2	8.3	11.4

9. Fit the following set of data through the relation $y = b\exp(ax)$ to get the values of a and b.

$x =$	1.2	1.9	3.5	5.2
$y =$	2.13	1.98	1.69	1.43

10. With the help of the Lagrange's interpolation formula obtain the value of y corresponding to $x = 3$ from the following data.

$x =$	2	4	5	8
$y =$	7	11	13	19

10. Write short notes on the following:

    (i) Graphical interpolation
    (ii) Forward difference scheme
    (iii) Backward difference scheme
    (iv) Central difference scheme

(v) Gauss' forward interpolation

(vi) Gauss' backward interpolation

(vii) Lagrange's interpolation

V. Hypotheses Testing

The tests may be classified into two categories: (i) parametric and (ii) non-parametric. Parametric tests are based on some distribution laws whereas non-parametric tests have no distribution laws. Here, we shall discuss about parametric tests. Non-parametric tests will be discussed in the following chapters.

Hypothesis behaves like a principal instrument in research. Quite often, its main function is to suggest new experiments and observations. In fact many experiments are carried out with the aim of testing hypotheses. Decision makers generally test hypotheses on the basis of available information and then take decisions on the basis of this testing. In case, the knowledge about population parameter(s) is not available, hypothesis testing is often used for deciding whether a sample data offer such support for a hypothesis that generalization can be made. Hence, hypothesis testing enables us to make probability statements about population parameter(s). The hypotheses may not be proved in absolute, but in practice it is accepted when it has withstood a critical testing. Before explaining how hypotheses are tested through different tests meant for the purpose, it will be appropriate to explain clearly the meaning of a hypothesis and related concepts for better understanding of hypotheses testing techniques.

1. Hypothesis

When one talks about a hypothesis, a common person may think it as an assumption or supposition to be proved or disproved on the basis of arguments. But, for a researcher, hypothesis is a formal question that one intends to resolve. Quite often a research hypothesis is a predictive statement, capable of being tested on the basis of scientific methods, which relates an independent variable to some dependent variable. For example, let us consider statements as the following ones:

"Students who are given private coaching in general show a better performance in examinations as compared to those who are not given private coaching."

"Performance of bike A is as good as that of bike B"

Such hypothesis can be objectively verified and tested. Thus, we may conclude that a hypothesis states what we are looking for and it is a proposition which can be put to a test for determining its validity.

1.1 Characteristics of a hypothesis

A hypothesis must have the following characteristics.

(i) Hypothesis should be clear and precise. When the hypothesis is not clear and precise, the conclusions drawn on its basis cannot be taken as reliable.

(ii) It should be possible to make a test for hypothesis. However, a researcher may do some prior study in order to make the hypothesis testable one. A hypothesis is testable when other deductions can be made from it which, in turn, can be proved or disproved on the basis of observations.

(iii) Hypothesis should be formulated in the form of a relationship between variables, when it is a relational hypothesis.

(iv) Hypothesis should be limited within the scope and must be specific. It is well known that narrower hypotheses are generally more testable and one can develop such hypotheses.

(v) Hypothesis should be stated as far as possible in most simple terms, so that the same is easily understandable by all concerned. However, it is obvious that simplicity of hypothesis has nothing to do with its significance.

(vi) Hypothesis should be consistent with most known facts. That is, it should be consistent with the established facts.

(vii) Hypothesis should be amenable to make testing with a reasonable time. That is, one should not go for, even an excellent hypothesis, when the same cannot be tested in a reasonable time one can spend for investigation.

(viii) Hypothesis must explain the facts that gave rise to the need for explanation. That is, by using the hypothesis along with other known and accepted generalizations, one should be able to deduce the original problem condition.

2. Basic concepts regarding hypothesis testing

Let us first discuss some basic concepts regarding hypotheses testing.

2.1 Null hypothesis and alternate hypothesis

In a hypothesis testing, there are two statistical hypotheses: (i) Null hypothesis denoted by H_0, and (ii) Alternate hypothesis denoted by H_A. The null hypothesis is the hypothesis to be tested. It is either rejected or not rejected. (Some authors consider the word 'accepted' in place of 'not rejected'. However, because of significance level, it appears appropriate to use the word 'not rejected'.) When the H_0 is rejected, the question raised may be acceptable. On the other side, when H_0 is not rejected, the question is not acceptable.

In establishing a hypothesis, an indication of equality (either $=$, $\leq$, or $\geq$) must appear in the null hypothesis. Suppose, for example, that we want to answer the question: Can we consider that a certain population mean is not μ_0? The null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : \mu = \mu_0 \qquad \text{and} \qquad H_A : \mu \neq \mu_0$$

Here, the alternate hypothesis is two sided, one side is $\mu < \mu_0$ and other side is $\mu > \mu_0$. When we want to know if we can conclude that the population mean is greater than μ_0. Our hypotheses for this situation are

$$H_0 : \mu \leq \mu_0 \qquad \text{and} \qquad H_A : \mu > \mu_0$$

Here, the alternate hypothesis is one sided. When we want to know if we can conclude that the population mean is less than μ_0. Our hypotheses for this situation are

$$H_0 : \mu \geq \mu_0 \qquad \text{and} \qquad H_A : \mu < \mu_0$$

Here also, the alternate hypothesis is one sided. In summary, we may state the following rules for deciding what statement goes in the null hypothesis and what statement goes in the alternate hypothesis:

(a) What we hope or expect to be able to conclude as a result of the test usually should be placed in the alternate hypothesis.

(b) The null hypothesis should contain a statement of equality. That is, for the null hypothesis we should have either $=$, $\leq$, or $\geq$. Hence, the hypotheses are tested for only three situations for alternate hypothesis: (i) not equal to, (ii) greater than, and (iii) smaller than.

(c) The null hypothesis is the hypothesis that is tested.

(d) The null and alternate hypotheses are complementary to each other. That is both of them together exhaust all possibilities regarding the values that the hypothesized parameter can assume.

It should be pointed out that the hypothesis testing merely indicates whether the hypothesis is supported or is not supported by the available data. When we fail to reject the null hypothesis, we do not say that it is true, but that it may be true. When we speak of accepting a null hypothesis, we do not wish to convey the idea that accepting implies proof.

Exercise 1: We want to investigate if the average income of people in India is not 600 rupees. Write the null hypothesis H_0 and alternate hypothesis H_A for the investigation.

Solution: The null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : \mu = 600 \qquad \text{and} \qquad H_A : \mu \neq 600$$

Exercise 2: We want to study if the average age of people in India is more than 40 years. Write the null hypothesis H_0 and alternate hypothesis H_A for the study.

Solution: The null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : \mu \leq 40 \qquad \text{and} \qquad H_A : \mu > 40$$

2.2 Significance level

For a normal distribution function

$$f(z) = \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-z^2/2}$$

the decision as to which values of z lie in the rejection region and which lie in the non-rejection region is made on the basis of the desired level of significance, designated by α. For one-side test (discussed later on), the probability of rejection is α and probability of non-rejection is $(1-\alpha)$. This value of α represents area on left or on right end of the distribution curve, depending on the hypotheses. This area is the rejection region and the remaining area under the distribution curve is non-rejection region. For two-sides test (discussed later on) also the probability of rejection is α and the probability of non-rejection is $(1-\alpha)$. In this case, the probability of rejection α is divided into two equal parts. Now, $\alpha/2$ is the area at each of the left and right ends of the distribution curve. The remaining area between these two rejection regions is non-rejection region.

Each of these areas are the rejection regions. Finally, when the calculated value of test-statistic corresponds to a rejection region, the null hypothesis is rejected. On the other hand, when the calculated value of test-statistic corresponds to a non-rejection region, the null hypothesis is not rejected.

2.3 Test statistic

Test statistic for testing null hypothesis is a parameter which may be computed from given data. It serves as a decision maker, since the decision to reject or not to reject the null hypothesis depends on the value of the test statistic. As an example, test statistic may be expressed as

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

where μ_0 is the value obtained from the hypotheses in the test. Here, $\overline{x}$ and σ^2 are, respectively, arithmetic mean and variance for the given n data.

Decision rule

When calculated value of test-statistic corresponds to a rejection region, the null hypothesis is rejected. On the other hand, when calculated value of test-statistic corresponds to a non-rejection region, the null hypothesis is not rejected.

2.4 Types of errors

When null hypothesis is tested, there are four possible situations:

1. The hypothesis is true but the test rejects it.
2. The hypothesis is true and the test does not reject it.
3. The hypothesis is false but the test does not reject it.
4. The hypothesis is false and the test rejects it.

When the null hypothesis is true, but it is rejected on the basis of the test, the error committed in this case is known as Type I error. The probability of committing a type I error is represented by α, also known as the significance level. When the null hypothesis is false, but it is not rejected on the basis of the test, the error committed in this case is known as Type II error. The probability of committing Type II error is represented by β. In other two cases: (i) the hypothesis is true and the test does not reject it, and (ii) the hypothesis is false and the test rejects it, the action is correct. Various possibilities for condition of null hypothesis and outcome of test are summarized in the following table. Such table is generally referred to as the 'confusion matrix'.

		Condition of null hypothesis	
		True	False
Possible Action	Fail to reject H_0	Correct action	Type II error
	Reject H_0	Type I error	Correct action

2.5 Two-sides test

In a two-sides test (also called two-tails test), null hypothesis is rejected when, say, the sample mean is higher than or lower than the value obtained from the significance level. Such a test is appropriate when null hypothesis is some specified value and alternate hypothesis is a value not equal to the specified value of the null hypothesis. Symbolically, a two-sides test is appropriate when we have $H_0 : \mu = \mu_0$ and $H_A : \mu \neq \mu_0$, where μ_0 is some specified value. Hence, in a two-sides test, there are two rejection regions, one on each tail of the curve (Figure 1). Mathematically, it can be expressed as

Non-rejection region: $-z_0 < z < z_0$

Rejection regions: $z < -z_0$ or $z > z_0$

The value of z_0 is obtained on the basis of the significance level. For example, when the significance level is 5%, the probability of rejection area is 0.05 and of non-rejection area is 0.95. Suppose a two-sides test is to be applied. For a two-sides test, the rejection area is divided equally on both tails. Thus, the rejection area on each tail is 0.025. From Table 1 in Appendix, this value 0.025 of area corresponds to $z = -1.96$ and the area $0.975 (= 1 - 0.025)$ corresponds to $z = 1.96$. The null hypothesis is rejected when computed value of test static is either large than 1.96 or smaller than -1.96. When the test static lies in between -1.96 and 1.96, the null hypothesis is not rejected.

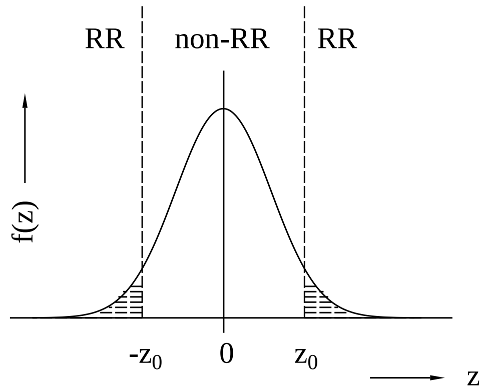


Figure 1: Rejection regions (denoted by RR) and non-rejection region (denoted by non-RR) in a two-sides test. The rejection regions are at the left as well as right tails.

2.6 One-side tests

There are situations when one-side test (also known as one-tail test) is appropriate rather than two-sides test. A one-side test would be used when we want to test, say, whether calculated value of given parameter is lower than or higher than some hypothesized (given) value for the parameter.

(i) Left-side test

When we have the situation $H_0 : \mu \geq \mu_0$ and $H_A : \mu < \mu_0$, then we are interested in what is known as left-side test. Here, μ_0 is the hypothesized value. Thus, there is only one rejection region, which is at the left tail as shown in Figure 2. Mathematically, it can be expressed as

Non-rejection region: $z > -z_0$

Rejection region: $z < -z_0$

The value of z_0 is obtained on the basis of the significance level. For example, when the significance level is 5%, the probability of the rejection area is 0.05 and non-rejection area is 0.95. Suppose, one-side test is to be applied. From Table 1 in appendix, this value 0.05 of area corresponds to $z = -1.645$. The null hypothesis is rejected when computed value of test static is less than -1.645. When the test static is larger than -1.645, the null hypothesis is not rejected.

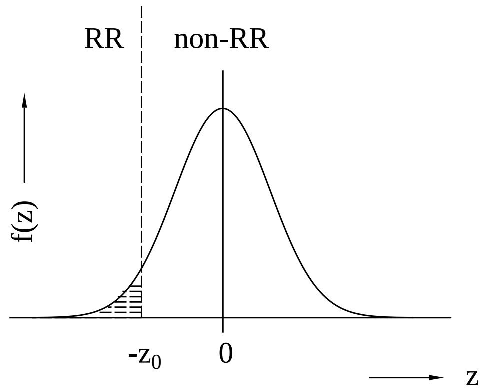


Figure 2: Rejection region (denoted by RR) and non-rejection region (denoted by non-RR) in one-side test. The rejection region is at the left tail.

(ii) Right-side test

When we have $H_0 : \mu \leq \mu_0$ and $H_A : \mu > \mu_0$, then we are interested in what is known as right-side test. Thus, there is only one rejection region, which is at the right tail as shown in Figure 3. Mathematically, it can be expressed as

Non-rejection region: $z < z_0$

Rejection region: $z > z_0$

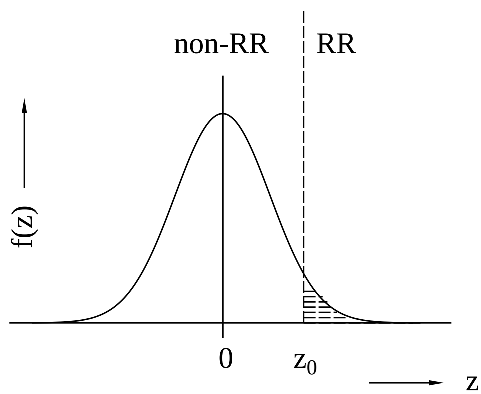


Figure 3: Rejection region (denoted by RR) and non-rejection region (denoted by non-RR) in one-side test. The rejection region is at the right tail.

The value of z_0 is obtained on the basis of the significance level. For example, when the significance level is 5%, the probability of the rejection area is 0.05 and non-rejection area is be 0.95. Suppose, we want one-side test is to be applied. The value 0.95 $(= 1 - 0.05)$ of area corresponds to $z = 1.645$. The null hypothesis is rejected when computed value of test static is larger than 1.645. When the test static is less than 1.645, the null hypothesis is not rejected.

3. Procedure for hypothesis testing

For data collected by a researcher, to test null hypothesis means to tell whether or not the hypothesis is valid. In the hypothesis testing, the main question is: whether to accept null hypothesis or not to accept null hypothesis. Procedure for hypothesis testing refers to all those steps which we undertake for making a choice between rejection and non-rejection of null hypothesis. Various steps involved in the hypothesis testing may be as the following.

(i) Formulation of statement: This step consists of making a formal statement of null hypothesis H_0 and alternate hypothesis H_A. That is, the hypotheses should be stated clearly, considering the nature of research problem. For example, we want to test if the mean of given data is greater than μ_0, the hypotheses are as the following:

$$H_0 : \mu \leq \mu_0 \qquad \text{and} \qquad H_A : \mu > \mu_0$$

Other possible situations for H_0 and H_A may be

$$H_0 : \mu \geq \mu_0 \qquad \text{and} \qquad H_A : \mu < \mu_0$$

and

$$H_0 : \mu = \mu_0 \qquad \text{and} \qquad H_A : \mu \neq \mu_0$$

The formulation of hypothesis is an important step which must be accomplished with due care in accordance with the object and nature of problem under consideration. It also indicates whether we should use one-tail test or two-tails test. When H_A is of the type 'greater than' or of the type 'smaller than', we use one-tail test, but when H_A is of the type 'not equal to', we use two-tails test.

(ii) Choice of significance level: Null hypothesis is tested on a pre-determined level of significance, denoted by α, which needs to be specified before calculating the test parameter, called test static. It is also known as Type I error. For example, we can consider 5% significance level. It show that under the normal distribution curve, rejection probability (area) is 0.05 and non-rejection probability (area) is 0.95.

(iii) Decision for a law of distribution: After deciding the level of significance, we determine an appropriate sampling distribution. For example, we can choose a normal distribution.

(iv) Computation of statistical parameters for a random sample: We select a random sample and calculate statistical parameters. For example, we calculate arithmetic mean and variance for the data in the sample.

(v) Computation of test-parameter: From the statistical parameters for the data in the sample, we calculate test-parameter, such as test static.

(vi) Making a decision: This step consists of comparing the calculated test static with the value obtained corresponding to α, the significance level. It helps in making a decision whether null hypothesis is rejected or not rejected.

Exercise 3: For a sample of marks obtained by 20 students with arithmetic mean 50 and variance 22, we want to investigate if the average marks obtained by the students is not 52. Considering the significance level $\alpha = 0.0702$, decide if the null hypothesis will be rejected or not. Also calculate p-value for the conclusion.

Solution: For the case stated here, the null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : \mu = 52 \qquad \text{and} \qquad H_A : \mu \neq 52$$

This is case of two-tails test. For two-tails test and significance level $\alpha = 0.0702$, we have $\alpha/2 = 0.0351$. From Table 1 in Appendix, the z-values corresponding to 0.0351 and $(1 - 0.0351 =)0.9649$ are -1.81 and $+1.81$, respectively. Thus, non-rejection region lies between $z = -1.81$ and $z = +1.81$. The rejection regions lie between $z = -\infty$ and $z = -1.81$, and between $z = +1.81$ and $z = \infty$. The test statistic is

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{50 - 52}{\sqrt{22/20}} = -1.91$$

Since the value -1.91 lies in the rejection region, the null hypothesis is rejected.

For $z = -1.91$ and $z = +1.91$, the values from Table 1 in Appendix are 0.0281 and 0.9719, respectively. Now, $1 - 0.9719 = 0.0281$. Thus, the p-value is $0.0281 + 0.0281 = 0.0562$.

Exercise 4: Considering the significance level $\alpha = 0.0351$, decide if the null hypothesis will be rejected or not. For a sample of marks obtained by 18 students with arithmetic mean 47 and variance 20, we want to investigate if the average marks obtained by students is more than 50.

Solution: For the case stated here, the null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : \mu \leq 50 \qquad \text{and} \qquad H_A : \mu > 50$$

This is case of one-tail test on right side. For one-tail test, we have significance level $\alpha = 0.0351$. From Table 1 in Appendix, the z-value corresponding to $1 - 0.0351 = 0.9649$ is 1.81. The rejection region lies between $z = +1.81$ and $z = \infty$. The non-rejection region lies between $z = -\infty$ and $z = +1.81$. The test statistic is

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{47 - 50}{\sqrt{20/18}} = -2.85$$

Since the value -2.85 lies in the non-rejection region, the null hypothesis is not rejected.

Exercise 5: Considering the significance level $\alpha = 0.0375$, decide if the null hypothesis will be rejected or not. We have data for ages of 22 people with arithmetic mean 43 and variance 24, and we want to investigate if the average age of the people in a village is less than 45.

Solution: For the case stated here, the null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : \mu \geq 45 \qquad \text{and} \qquad H_A : \mu < 45$$

This is case of one-tail test on left side. For one-side test, we have significance level $\alpha = 0.0375$. From Table 1 in Appendix, the z-value corresponding to 0.0375 is -1.78. The non-rejection region is between $z = -1.78$ and $z = +\infty$. The rejection regions lies between $z = -\infty$ and $z = -1.78$. The test statistic is

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{43 - 45}{\sqrt{24/22}} = -1.91$$

Since the value -1.91 lies in the rejection region, the null hypothesis is rejected.

4. Power of the test

As mentioned earlier, there are two types of errors: Type I error and Type II error. The probability of Type I error is denoted as α (also called the significance level) and the probability of Type II error is denoted as β. Generally, the significance level for a test is assigned in advance and once it is decided, one cannot do any thing else about it.

We all know that hypothesis test cannot be fool proof; sometimes the test does not reject H_0 when it happens to be false. This way a Type II error is committed. Since Type II error can be controlled, one would like that β is as small as possible. Alternatively, one would like that $(1-\beta)$ is as large as possible. When $(1-\beta)$ is close to 1.0, we can infer that the test is working quite well, meaning thereby that the test is rejecting H_0 when H_0 is not true. On the other side, when $(1-\beta)$ is close to 0.0, we can infer that the test is working poorly, meaning thereby that the test is not rejecting H_0 when H_0 is not true. Thus, the value of $(1-\beta)$ is a measure of how well the test is working. It is technically described as the 'power of the test'.

When we plot the values of $(1-\beta)$ for each possible value of the population parameter (say μ, the true population mean) for which the H_0 is not true, the resulting curve is known as the power curve for the given test. Thus, the power curve of a hypothesis test is the curve that shows the conditional probability of rejecting H_0 as a function of the population parameter and size of the sample. The function defining this curve is known as the power function. In other words, the power function of a test is that function defined for all values of the parameter(s) which yields the probability that H_0 is rejected. Value of the power function at a specific parameter point is called the power of the test at that point.

Exercise 6: The null hypothesis H_0 and the alternate hypothesis H_A for an investigation are

$$H_0 : \mu = 55 \qquad \text{and} \qquad H_A : \mu \neq 55$$

Considering the significance level $\alpha = 0.075$ and normal distribution, when the H_0 is false, calculate the power of test for $\mu_0 = 56.3$. For a given sample of 65 data, the variance is 30.

Solution: This is case of two-sides tail. For normal distribution, for the significance level $\alpha = 0.075$, we have $\alpha/2 = 0.0375$. From Table 1 in Appendix, corresponding to 0.0375, we have $z = -1.78$. The lower and upper limits are

$$\overline{x}_L = 55 - 1.78\ \sqrt{30/65} = 53.79$$

and

$$\overline{x}_U = 55 + 1.78\ \sqrt{30/65} = 56.21$$

We are given that the H_0 is false, then for $\mu_0 = 56.3$, the limits for z are

$$z_L = \frac{53.79 - 56.3}{\sqrt{30/65}} = -3.69 \qquad \text{and} \qquad z_U = \frac{56.21 - 56.3}{\sqrt{30/65}} = -0.132$$

For these z-values, form Table 1 in Appendix, we get $A_L \approx 0.0$ and $A_U \approx 0.4483$. Thus, $\beta = A_U - A_L = 0.4483$. Thus, the power of test is $1 - 0.4483 = 0.5517$.

5. Power curve for two-sides test

For having power curve for two-sides test, one proceeds as the following. Let us consider a set of n data with variance σ^2. Suppose, we want to make a test with significance level α and hypotheses

$$H_0 : \mu = \mu_0 \qquad \text{and} \qquad H_A : \mu \neq \mu_0$$

For two-sides test, significance level is divided into two equal parts, so that each part is $\alpha/2$. From Table 1 in Appendix, corresponding to $\alpha/2$, we get a value z_0. Now, corresponding to μ_0, we calculate lower and upper limits as the following.

$$\overline{x}_L = \mu_0 - |z_0|\ \frac{\sigma}{\sqrt{n}} \qquad \text{and} \qquad \overline{x}_U = \mu_0 + |z_0|\ \frac{\sigma}{\sqrt{n}}$$

Now, we consider various values for μ^* which lie on the two sides of μ_0. For each value of μ^*, we calculate z_L and z_U as the following.

$$z_L = \frac{\overline{x}_L - \mu^*}{\sigma/\sqrt{n}} \qquad \text{and} \qquad z_U = \frac{\overline{x}_U - \mu^*}{\sigma/\sqrt{n}}$$

From Table 1 in Appendix, corresponding to z_L and z_U, we obtain the values A_L and A_U, respectively. Then, β is defined as $\beta = A_U - A_L$ and the power of test is $(1 - \beta)$.

Exercise 7: With significance level 0.0784, for a set of 30 data having variance 12.96 and hypotheses

$$H_0 : \mu = 60 \qquad \text{and} \qquad H_A : \mu \neq 60$$

find out the power curve.

Solution: The hypotheses in the present case are as the following.

$$H_0 : \mu = 60 \qquad \text{and} \qquad H_A : \mu \neq 60$$

This is case of two sides test with significance level $\alpha = 0.0784$. Thus, each side has $\alpha/2 = 0.0392$. From Table 1 in Appendix, corresponding to 0.0392, we have $z = -1.76$. For $\mu_0 = 60$, we calculate lower and upper limits as the following.

$$\overline{x}_L = \mu_0 - |z| \frac{\sigma}{\sqrt{n}} = 60 - 1.76 \sqrt{\frac{12.96}{30}} = 58.84$$

and

$$\overline{x}_U = \mu_0 + |z| \frac{\sigma}{\sqrt{n}} = 60 + 1.76 \sqrt{\frac{12.96}{30}} = 61.16$$

We consider various values for μ^* which lie on the two sides of $\mu_0 = 60$. We accounted for μ^* from 57.0 to 63.0 with increment of 0.5. For each value of μ^*, we calculate z_L and z_U as the following.

$$z_L = \frac{\overline{x}_L - \mu^*}{\sigma/\sqrt{n}} = \frac{58.84 - \mu^*}{\sqrt{12.96/30}}$$

and

$$z_U = \frac{\overline{x}_U - \mu^*}{\sigma/\sqrt{n}} = \frac{61.16 - \mu^*}{\sqrt{12.96/30}}$$

The values of z_L and z_U corresponding to μ^* are given in columns 2 and 3 of the following table. From Table 1 in Appendix, values A_L and A_U corresponding to z_L and z_U are obtained and are given in columns 4 and 5 of the table. Values of $\beta = A_U - A_L$ and $(1 - \beta)$ are given columns 6 and 7 of the table. The values of $1 - \beta$ for various values of μ^* are plotted in Figure 4.

μ^*	z_L	z_U	A_L	A_U	β	$1-\beta$
57.0	2.80	6.33	1.00	1.00	0.0026	0.9974
57.5	2.04	5.57	0.98	1.00	0.0207	0.9793
58.0	1.28	4.81	0.90	1.00	0.1003	0.8997
58.5	0.52	4.05	0.70	1.00	0.3015	0.6985
59.0	-0.24	3.29	0.41	1.00	0.5943	0.4057
59.5	-1.00	2.53	0.16	0.99	0.8356	0.1644
60.0	-1.76	1.76	0.04	0.96	0.9216	0.0784
60.5	-2.53	1.00	0.01	0.84	0.8356	0.1644
61.0	-3.29	0.24	0.00	0.59	0.5943	0.4057
61.5	-4.05	-0.52	0.00	0.30	0.3015	0.6985
62.0	-4.81	-1.28	0.00	0.10	0.1003	0.8997
62.5	-5.57	-2.04	0.00	0.02	0.0207	0.9793
63.0	-6.33	-2.80	0.00	0.00	0.0026	0.9974

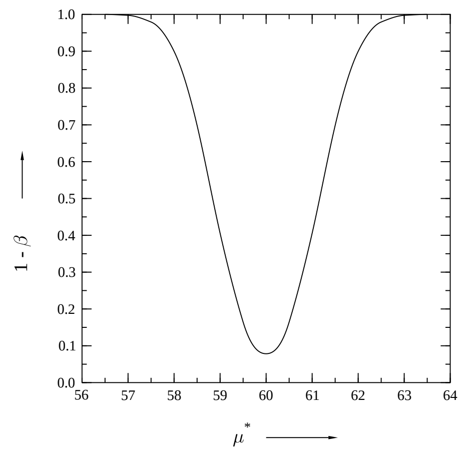


Figure 4: Variation of $1-\beta$ versus μ^*.

Exercise 8: With significance level 0.05, for a set of 100 data having variance 12.96 and hypotheses

$$H_0 : \mu = 17.5 \qquad \text{and} \qquad H_A : \mu \neq 17.5$$

find out the power curve.

Solution: The hypotheses in the present case are as the following.

$$H_0 : \mu = 17.5 \qquad \text{and} \qquad H_A : \mu \neq 17.5$$

This is case of two sides test with significance level $\alpha = 0.0784$. Thus, each side has $\alpha/2 = 0.025$. From Table 1 in Appendix, corresponding to 0.025, we have $z = -1.96$. For $\mu_0 = 17.5$, we calculate lower and upper limits as the following.

$$\overline{x}_L = \mu_0 - |z| \, \frac{\sigma}{\sqrt{n}} = 17.5 - 1.96 \, \sqrt{\frac{12.96}{100}} = 16.79$$

and

$$\overline{x}_U = \mu_0 + |z| \, \frac{\sigma}{\sqrt{n}} = 17.5 + 1.96 \, \sqrt{\frac{12.96}{100}} = 18.21$$

We consider various values for μ^* which lie on the two sides of $\mu_0 = 17.5$. We accounted for μ^* from 16.0 to 19.0 with increment of 0.5. For each value of μ^*, we calculate z_L and z_U as the following.

$$z_L = \frac{\overline{x}_L - \mu^*}{\sigma/\sqrt{n}} = \frac{16.79 - \mu^*}{\sqrt{12.96/100}}$$

and

$$z_U = \frac{\overline{x}_U - \mu^*}{\sigma/\sqrt{n}} = \frac{18.21 - \mu^*}{\sqrt{12.96/100}}$$

The values of z_L and z_U corresponding to μ^* are given in columns 2 and 3 of the following table. From Table 1 in Appendix, values A_L and A_U corresponding to z_L and z_U are obtained and are given in columns 4 and 5 of the table. Values of $\beta = A_U - A_L$ and $(1 - \beta)$ are given columns 6 and 7 of the table. The values of $1 - \beta$ for various values of μ^* are plotted in Figure 5.

μ^*	z_L	z_U	A_L	A_U	β	$1 - \beta$
16.0	2.19	6.14	0.99	1.00	0.0143	0.9857
16.5	0.81	4.75	0.79	1.00	0.2090	0.7910
17.0	-0.58	3.36	0.28	1.00	0.7186	0.2814
17.5	-1.97	1.97	0.02	0.98	0.9512	0.0488
18.0	-3.36	0.58	0.00	0.72	0.7186	0.2814
18.5	-4.75	-0.81	0.00	0.21	0.2090	0.7910
19.0	-6.14	-2.19	0.00	0.01	0.0143	0.9857

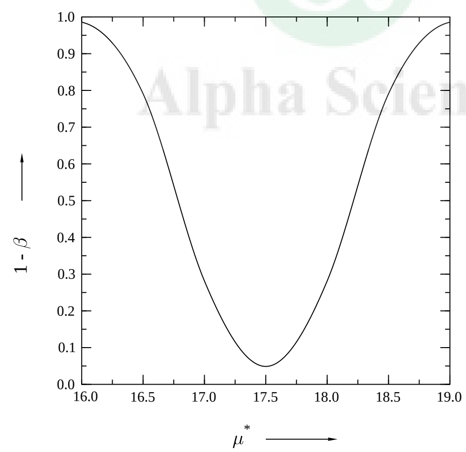


Figure 5: Variation of $1 - \beta$ versus μ^*.

6. Power curve for one-side tests

For one-side test, there are two situations: (i) left-side test and (ii) right-side test. We shall discuss these two cases separately.

6.1 Power curve for left-side test

For having power curve for left-side test, one proceeds as the following. Let us consider a set of n data with variance σ^2. Suppose, we want to make a test with significance level α and hypotheses

$$H_0 : \mu \geq \mu_0 \qquad \text{and} \qquad H_A : \mu < \mu_0$$

For such test, significance level α is the area in the left tail. From Table 1 in Appendix, corresponding to α, we get a value z_0. Now, corresponding to μ_0, we calculate lower limit as the following.

$$\overline{x}_L = \mu_0 - |z_0| \frac{\sigma}{\sqrt{n}}$$

In fact, z_0 is a negative value. Here, the upper limit $\overline{x}_U$ is infinite. Now, we consider various values for μ^* which lie on the lower side of μ_0. For each value of μ^*, we calculate z_L as the following.

$$z_L = \frac{\overline{x}_L - \mu^*}{\sigma/\sqrt{n}}$$

Now, corresponding to $\overline{x}_U$, the value z_U is infinite. From Table 1 in Appendix, corresponding to z_L and z_U, we obtain the valuesA_L and 1.0, respectively. Then, β is defined as $\beta = 1 - A_L$ and the power of test is $(1 - \beta)$.

Exercise 9: With significance level 0.0202, for a set of 25 data having variance 324 and hypotheses

$$H_0 : \mu \geq 60 \qquad \text{and} \qquad H_A : \mu < 60$$

find out the power curve.

Solution: The hypotheses in the present case are as the following.

$$H_0 : \mu \geq 60 \qquad \text{and} \qquad H_A : \mu < 60$$

This is case of one side test with significance level $\alpha = 0.0202$. From Table 1 in Appendix, corresponding to 0.0202, we have $z = -2.05$. For $\mu_0 = 60$, we calculate lower limit as the following.

$$\overline{x}_L = \mu_0 - |z| \frac{\sigma}{\sqrt{n}} = 60 - 2.05 \sqrt{\frac{324}{25}} = 52.62$$

The upper limit $\overline{x}_U$ is infinite. Now, we consider various values for μ^* which lie on the lower side of 60. We accounted for μ^* from 45 to 60 with increment of 1. For each value of μ^*, we calculate z_L is expressed as

$$z_L = \frac{\overline{x}_L - \mu}{\sigma/\sqrt{n}} = \frac{52.62 - \mu^*}{\sqrt{324/25}}$$

Now, corresponding to $\overline{x}_U$, the value z_U is infinite. Values of z_L are given in column 2 of Table 4. From Table 1 in Appendix, corresponding to z_L and z_U, we obtain the values A_L and 1.00, respectively. Values of A_L are given in column 3 of the table. Values of $\beta = 1 - A_L$ and $(1-\beta)$ are given columns 4 and 5 of the table. The values of $1-\beta$ for various values of μ^* are plotted in Figure 6.

μ^*	z_L	A_L	β	$1-\beta$
45	2.12	0.98	0.0170	0.9830
46	1.84	0.97	0.0329	0.9671
47	1.56	0.94	0.0594	0.9406
48	1.28	0.90	0.1003	0.8997
49	1.01	0.84	0.1562	0.8438
50	0.73	0.77	0.2327	0.7673
51	0.45	0.67	0.3264	0.6736
52	0.17	0.57	0.4325	0.5675
53	-0.11	0.46	0.5438	0.4562
54	-0.38	0.35	0.6480	0.3520
55	-0.66	0.25	0.7454	0.2546
56	-0.94	0.17	0.8264	0.1736
57	-1.22	0.11	0.8888	0.1112
58	-1.49	0.07	0.9319	0.0681
59	-1.77	0.04	0.9616	0.0384
60	-2.05	0.02	0.9798	0.0202

6.2 Power curve for right-side test

For having power curve for right-side test, one proceeds as the following. Let us consider a set of n data with variance σ^2. Suppose, we want to make a test with significance level α and hypotheses

$$H_0 : \mu \leq \mu_0 \qquad \text{and} \qquad H_A : \mu > \mu_0$$

For such test, significance level α is the area in the right tail. From Table 1 in Appendix, corresponding to $(1-\alpha)$, we get a value z_0. Now, corresponding to μ_0, we calculate upper limit as the following.

$$\overline{x}_U = \mu_0 + |z_0| \, \frac{\sigma}{\sqrt{n}}$$

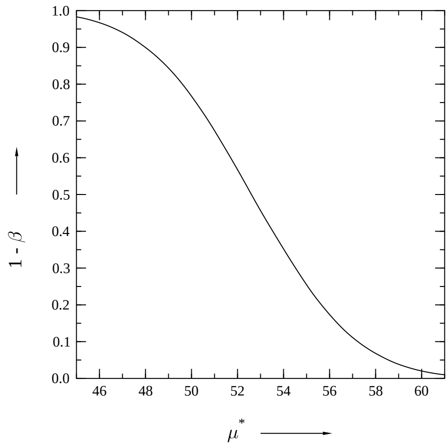


Figure 6: Variation of $1 - \beta$ versus μ^*.

In fact, z_0 is a positive value. The lower limit $\overline{x}_L$ is $-\infty$. Now, we consider various values for μ^* which lie on the higher side of μ_0. For each value of μ^*, we calculate z_U is expressed as

$$z_U = \frac{\overline{x}_U - \mu^*}{\sigma/\sqrt{n}}$$

Now, corresponding to $\overline{x}_L$, the value z_L is $-\infty$. From Table 1 in Appendix, corresponding to z_L and z_U, we obtain the values 0.0 and A_U, respectively. Then, β is defined as $\beta = A_U - 0$ and the power of test is $(1 - \beta)$.

Exercise 10: With significance level 0.0202, for a set of 25 data having variance 324 and hypotheses

$$H_0 : \mu \leq 60 \qquad \text{and} \qquad H_A : \mu > 60$$

find out the power curve.

Solution: The hypotheses in the present case are as the following.

$$H_0 : \mu \leq 60 \qquad \text{and} \qquad H_A : \mu > 60$$

This is case of one side test with significance level $\alpha = 0.0202$. From Table 1 in Appendix, corresponding to $(1 - 0.0202)$, we have $z = 2.05$. For $\mu_0 = 60$, we calculate upper limit as the following.

$$\overline{x}_U = \mu_0 + |z| \, \frac{\sigma}{\sqrt{n}} = 60 + 2.05 \, \sqrt{\frac{324}{25}} = 67.38$$

Here, the lower limit $\overline{x}_L$ is $-\infty$. Now, we consider various values for μ^* which lie on the higher side of 60. We accounted for μ^* from 60 to 75 with increment of 1. For each value of μ^*, we calculate z_U as the following.

$$z_U = \frac{\overline{x}_U - \mu}{\sigma/\sqrt{n}} = \frac{67.38 - \mu^*}{\sqrt{324/25}}$$

Now, corresponding to $\overline{x}_L$, the value z_L is $-\infty$. Values of z_U are given in column 2 of Table 5. From Table 1 in Appendix, corresponding to z_L and z_U, we obtain the values 0.0 and A_U, respectively. Values of A_U are given in column 3 of the table. Values of $\beta = A_U - 0.0$ and $(1-\beta)$ are given columns 4 and 5 of the table. The values of $1-\beta$ for various values of μ^* are plotted in Figure 7.

μ^*	z_U	A_U	β	$1-\beta$
60.0	2.05	0.98	0.9798	0.0202
61.0	1.77	0.96	0.9616	0.0384
62.0	1.49	0.93	0.9319	0.0681
63.0	1.22	0.89	0.8888	0.1112
64.0	0.94	0.83	0.8264	0.1736
65.0	0.66	0.75	0.7454	0.2546
66.0	0.38	0.65	0.6480	0.3520
67.0	0.11	0.54	0.5438	0.4562
68.0	-0.17	0.43	0.4325	0.5675
69.0	-0.45	0.33	0.3264	0.6736
70.0	-0.73	0.23	0.2327	0.7673
71.0	-1.01	0.16	0.1562	0.8438
72.0	-1.28	0.10	0.1003	0.8997
73.0	-1.56	0.06	0.0594	0.9406
74.0	-1.84	0.03	0.0329	0.9671
75.0	-2.12	0.02	0.0170	0.9830

7. Sample size to control type II error

Once we decide about significance level α for testing hypotheses, there is no way out to control Type I error. But, it is possible to check Type II error by controlling the sample size n. While discussing about power curve for a test, we have seen that type II error β was $\beta = A_U - A_L$ (for two-sides test), $\beta = 1 - A_L$ (for left-side test) and $\beta = A_U$ (for right-side test). For a given value of β, in case of one-side test (left or right), it is convenient to find out z_L (corresponding to A_L) or z_U (corresponding to A_U) for a given value of β and accordingly the sample size n may be

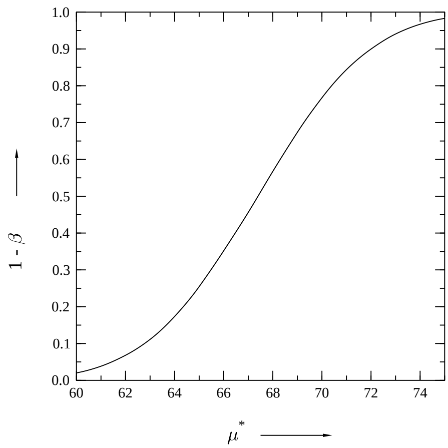


Figure 7: Variation of $1-\beta$ versus μ^*.

to controlled. Thus, here we shall discuss about sample size to control type II error for one-side tests.

7.1 For left-side test

Let us consider a set of data with variance σ^2. Suppose, we want to make a test with significance level α and hypotheses

$$H_0 : \mu \geq \mu_0 \qquad \text{and} \qquad H_A : \mu < \mu_0$$

For such test, significance level α is the area in the left tail. From Table 1 in Appendix, corresponding to α, we get a value z_0. Now, corresponding to μ_0, we calculate lower limit as the following.

$$\overline{x}_L = \mu_0 - |z_0| \, \frac{\sigma}{\sqrt{n}} \tag{5.1}$$

In fact, z_0 is a negative value. Now, we consider a value for μ^* which lies on the lower side of μ_0. For this value of μ^*, we calculate z_L as the following.

$$z_L = \frac{\overline{x}_L - \mu^*}{\sigma/\sqrt{n}} \tag{5.2}$$

From Table 1 in Appendix, corresponding to z_L, we obtain the values A_L and β is defined as $\beta = 1 - A_L$. Using equation (5.1) in (5.2), we get

$$n = \Big[\frac{(|z_0| + z_L)\sigma}{\mu_0 - \mu^*}\Big]^2$$

For a given value of β, we know A_L and corresponding to area A_L, we can find out z_L. z_0 is obtained corresponding to the area α, which is the significance level.

Exercise 11: The population standard deviation is 22 and the probability of type I error is set at 0.01. Suppose we want the probability of failing to reject $H_0(\beta)$ to be 0.05 when H_0 is false as the true mean is 54 rather than the hypothesized 62. How large the sample do we need in order to realize, simultaneously, the desired levels of α and β?

Solution: Here, we have $\mu_0 = 62$, $\mu^* = 54$, $\sigma = 22$, $\alpha = 0.01$ and $\beta = 0.05$. Thus, $A_L = 1 - \beta = 0.95$. Corresponding to area $\alpha = 0.01$, we have $z_0 = -2.33$ and corresponding to area $A_L = 0.95$, we have $z_1 = 1.645$. Hence, we have

$$n = \Big[\frac{(|z_0| + z_1)\sigma}{(\mu_0 - \mu^*)}\Big]^2 = \Big[\frac{(2.33 + 1.645)22}{(62 - 54)}\Big]^2 = 119.49$$

We would need a sample of 120 items to achieve the desired values of α and β when we take $\mu^* = 54$ as the alternative value of $\mu_0 = 62$.

7.2 For right-side test

Let us consider a set of data with variance σ^2. Suppose, we want to make a test with significance level α and hypotheses

$$H_0 : \mu \leq \mu_0 \qquad \text{and} \qquad H_A : \mu > \mu_0$$

For such test, significance level α is the area in the right tail. From Table 1 in Appendix, corresponding to $(1 - \alpha)$, we get a value z_0. Now, corresponding to μ_0, we calculate upper limit as the following.

$$\overline{x}_U = \mu_0 + |z_0| \, \frac{\sigma}{\sqrt{n}} \tag{5.3}$$

In fact, z_0 is a positive value. Now, we consider a value for μ^* which lies on the upper side of μ_0. For this value of μ^*, we calculate z_U as the following.

$$z_U = \frac{\overline{x}_U - \mu^*}{\sigma/\sqrt{n}} \tag{5.4}$$

From Table 1 in Appendix, corresponding to z_U, we obtain the values A_U and β is defined as $\beta = A_U$. Using equation (5.3) in (5.4), we get

$$n = \Big[\frac{(|z_0| - z_U)\sigma}{\mu^* - \mu_0}\Big]^2$$

For a given value of β, we know A_U and corresponding to area A_U, we can find out z_U. z_0 is obtained corresponding to the area α, which is the significance level.

Exercise 12: The population standard deviation is 22 and the probability of type I error is set at 0.01. Suppose we want the probability of failing to reject $H_0(\beta)$ to be 0.05 when H_0 is false as the true mean is 68 rather than the hypothesized 62. How large the sample do we need in order to realize, simultaneously, the desired levels of α and β?

Solution: Here, we have $\mu_0 = 62$, $\mu^* = 68$, $\sigma = 22$, $\alpha = 0.01$ and $\beta = 0.05$. Thus, $A_U = \beta = 0.05$. Corresponding to area $1 - \alpha = 0.99$, we have $z_0 = 2.33$ and corresponding to area $A_U = 0.05$, we have $z_1 = -1.645$. Hence, we have

$$n = \Big[\frac{(|z_0| + z_1)\sigma}{(\mu^* - \mu_1)}\Big]^2 = \Big[\frac{(2.33 - 1.645)22}{(68 - 62)}\Big]^2 = 6.3$$

We would need a sample of 7 items to achieve the desired values of α and β when we take $\mu^* = 68$ as the alternative value of $\mu_0 = 62$.

8. Two sample means

Frequently, we are interested in an investigation of two populations. For example, we may wish to know something about the difference between two population means; we may wish to know if it is reasonable to conclude that two population means are different. When we are able to conclude that the population means are different, we would like to know how much they differ. A knowledge of the sampling distribution of the difference between two means is useful in the investigations of this type.

8.1 Sampling from normally distributed populations

Suppose we have two populations, where one population (population 1) has gone through some special training, whereas the other population (population 2) has not gone through such training. The distribution of intelligence scores in each of the two populations is believed to be approximately normal distribution with a standard deviation σ. Now, suppose, that we take a sample of n individuals from each population and compute the mean intelligence scores for each sample with the results $\overline{x}_1$ and $\overline{x}_2$.

8.2 Construction of sampling distribution of $\overline{x}_1 - \overline{x}_2$

Although in practice, we do not attempt to construct the desired sampling distribution, we can however conceptualize the manner in which it could be done when the sampling is from finite populations. We shall begin by selecting from population 1 all possible samples of size 20 and computing the mean for each of the samples. For the population size N_1 of population 1 and sample size n_1, there would be ${}^{N_1}C_{n_1}$ samples. Here, $n_1 = 20$. Similarly, for the population size N_2 of population 2 and sample size n_2, there would be ${}^{N_2}C_{n_2}$ samples. Here, $n_2 = 20$. We shall then take all possible pairs of sample means, one from population 1 and other from population 2, and take the difference. Table 1 shows the results of this procedure.

Table 1. Difference between two sample means

Sample from pop 1	Sample from pop 2	Sample means pop 1	Sample means pop 2	Possible differences between means
n_{11}	n_{12}	$\overline{x}_{11}$	$\overline{x}_{12}$	$\overline{x}_{11} - \overline{x}_{12}$
n_{21}	n_{22}	$\overline{x}_{21}$	$\overline{x}_{22}$	$\overline{x}_{21} - \overline{x}_{22}$
n_{31}	n_{32}	$\overline{x}_{31}$	$\overline{x}_{32}$	$\overline{x}_{31} - \overline{x}_{32}$
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

8.3 Characteristics of sampling distribution of $\overline{x}_1 - \overline{x}_2$

When we plot the sample differences against their frequency of occurrence, we obtain a normal distribution with a mean equal to $(\mu_1 - \mu_2)$, the difference between two populations means, and a variance equal to $(\sigma_1^2/n_1) + (\sigma_2^2/n_2)$, as shown in Figure 8.

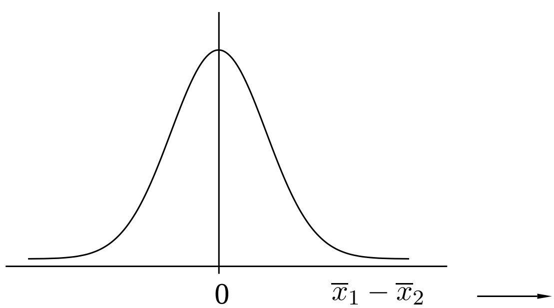


Figure 8: Distribution of $\overline{x}_1 - \overline{x}_2$ with center at $\mu_{\overline{x}_1 - \overline{x}_2} = \mu_1 - \mu_2 = 0$ and variance $\sigma_{\overline{x}_1 - \overline{x}_2} = (\sigma_1^2/n_1) + (\sigma_2^2/n_2)$.

8.4 Hypothesis testing

The hypothesis testing involving the difference between two population means is most frequently used to determine whether or not it is reasonable to conclude that the two population means μ_1 and μ_2 are unequal. In such case, the hypotheses which may be formulated are as the following.

1. $H_0 : (\mu_1 - \mu_2) = 0$ and $H_A : (\mu_1 - \mu_2) \neq 0$
2. $H_0 : (\mu_1 - \mu_2) \geq 0$ and $H_A : (\mu_1 - \mu_2) < 0$
3. $H_0 : (\mu_1 - \mu_2) \leq 0$ and $H_A : (\mu_1 - \mu_2) > 0$

It is possible to test the hypothesis that the difference is equal to, greater than or equal to, or less than or equal to some value other than zero. When each of two independent simple random samples have been drawn from a normally distributed population with a known variance, the test static for testing the null hypothesis of equal population means is

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

The rejection and non-rejection regions are decided on the basis of significance level α.

Exercise 13: For two samples, taken from two normally distributed populations, we have $n_1 = 30$, $n_2 = 35$, $\overline{x}_1 = 25$, $\overline{x}_2 = 27$, $\sigma_1^2 = 26$, $\sigma_2^2 = 28$. Using the significance level $\alpha = 0.05$, test hypotheses that population means are not equal.

Solution: For the case stated here, the null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : (\mu_1 - \mu_2) = 0 \qquad \text{and} \qquad H_A : (\mu_1 - \mu_2) \neq 0$$

This is case of two-tails test. For two-tails test and significance level $\alpha = 0.05$, we have $\alpha/2 = 0.025$. From Table 1 in Appendix, the z-values corresponding to 0.025 and $(1 - 0.025 =)0.975$ are -1.96 and $+1.96$, respectively. Thus, non-rejection region lies between $z = -1.96$ and $z = +1.96$. The rejection regions lie between $z = -\infty$ and $z = -1.96$, and between $z = +1.96$ and $z = \infty$. The test statistic is

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{(25 - 27) - 0}{\sqrt{(26/30) + (28/35)}} = -1.2$$

Since $z = -1.2$ lies in the rejection region, the null hypothesis is rejected.

9. Paired comparisons

In the preceding section for the differences between two population means, it was assumed that the samples were independent. Quite often, for example, we assess the effectiveness of a treatment or experimental procedure where we make use of related observations. For the hypothesized population mean difference μ_{d_0}, the hypotheses which may be formulated are as the following.

1. $H_0 : \mu_d = \mu_{d_0}$ and $H_A : \mu_d \neq \mu_{d_0}$
2. $H_0 : \mu_d \geq \mu_{d_0}$ and $H_A : \mu_d < \mu_{d_0}$
3. $H_0 : \mu_d \leq \mu_{d_0}$ and $H_A : \mu_d > \mu_{d_0}$

The test static for testing the null hypothesis is the Student's t static expressed as

$$t = \frac{\overline{d} - \mu_{d_0}}{\sqrt{s_d^2/n}}$$

where $\overline{d}$ is the sample mean difference, n the number of sample differences and s_d^2 the sample variance. The decision of rejection and non-rejection of null hypothesis is taken on the basis of significance level α and the degree of freedom $(n-1)$.

Exercise 14: For the pair of data given in the following table with significance level $\alpha = 0.05$ decide if the population mean difference is larger than zero.

Gallbladder % functions in patients before and after treatment

b	21	62.3	95	8.4	2.9	48	31	59	64	17.5
a	62.6	89.5	55.2	35.7	9.6	18.5	43.2	89.3	87.2	53.3

Solution: For the case stated here, the null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : \mu_d \leq 0 \qquad \text{and} \qquad H_A : \mu_d > 0$$

Let us first calculate the rejection and non-rejection regions. According to the significance level α and the degrees of freedom $n-1$, the critical value of t is obtained. For $n = 10$, the degrees of freedom $n - 1 = 9$. Now, with $\alpha = 0.05$ we have 95% confidence level and the critical value of t from Table 3 in Appendix is 1.8331. Now, the null hypothesis H_0 is rejected when computed t is larger than or equal to the critical value 1.8331. For calculation of test static, we prepare the following table.

a	b	$d_i = b - a$	$(d_i - \overline{d})$	$(d_i - \overline{d})^2$
21.0	62.6	41.6	28.100	789.610
62.3	89.5	27.2	13.700	187.690
95.0	55.2	-39.8	-53.300	2840.890
8.4	35.7	27.3	13.800	190.440
2.9	9.6	6.7	-6.800	46.240
48.0	18.5	-29.5	-43.000	1849.000
31.0	43.2	12.2	-1.300	1.690
59.0	89.3	30.3	16.800	282.240
64.0	87.2	23.2	9.700	94.090
17.5	53.3	35.8	22.300	497.290
	sum	135		6779.18

We have

$$\overline{d} = \frac{\sum d_i}{n} = \frac{135}{10} = 13.5$$

and

$$s_d^2 = \frac{\sum(d_i - \overline{d})^2}{n-1} = \frac{6779.18}{9} = 753.242$$

The test static is

$$t = \frac{\overline{d} - \mu_{d_0}}{\sqrt{s_d^2/n}} = \frac{13.5 - 0}{\sqrt{753.242/10}} = 1.555$$

Since 1.555 lies in the non-rejection region, null hypothesis H_0 cannot be rejected.

10. Single population proportion

Suppose a study is carried out on n samples and out of them x have shown the characteristic of interest. The population proportion is defined as

$$p = \frac{x}{n}$$

Testing of hypothesis about population proportion is carried out in a similar manner as for the means when the conditions necessary for using the normal curve are available. For the hypothesized population proportion p_0, possible hypotheses are as the following.

1. $H_0 : p = p_0$ and $H_A : p \neq p_0$
2. $H_0 : p \geq p_0$ and $H_A : p < p_0$

3. $H_0 : p \leq p_0$ and $H_A : p > p_0$

The test static is

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \qquad \text{where} \qquad \sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$$

The decision regarding rejection and non-rejection of null hypothesis is taken on the basis of significance level α.

Exercise 15: Data are obtained from the response of 301 individuals of which 24 possessed the characteristic of interest. For the significance level 0.05, decide if the population proportion is lager than 0.063.

Solution: For the case stated here, the null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : p \leq 0.063 \qquad \text{and} \qquad H_A : p > 0.063$$

Thus, we have

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.063(1-0.063)}{301}} = 0.014$$

This is case of one-tail test on right side. For one-tail test, we have significance level $\alpha = 0.05$. From Table 1 in Appendix, the z-value corresponding to $1 - 0.05 = 0.95$ is 1.645. The rejection region lies between $z = +1.645$ and $z = \infty$. The non-rejection region lies between $z = -\infty$ and $z = +1.645$. We have $\hat{p} = 24/301 = 0.080$. The test statistic is

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.080 - 0.063}{0.014} = 1.21$$

Since $z = 1.21$ lies in non-rejection region, the null hypothesis cannot be rejected.

11. Difference between two population proportions

Suppose a study is carried out on n_1 and n_2 samples and out of them x_1 and x_2, respectively, have shown the characteristic of interest. The population proportions are expressed as

$$p_1 = \frac{x_1}{n_1} \qquad \text{and} \qquad p_2 = \frac{x_2}{n_2}$$

Now, we define a parameter

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Testing of hypothesis about difference between two population proportions is carried out in a similar manner as for the means when the conditions necessary for using the normal curve are available. For the hypothesized difference between two population proportions p_M, possible hypotheses are as the following.

1. $H_0 : p_F = p_M$ and $H_A : p_F \neq p_M$
2. $H_0 : p_F \geq p_M$ and $H_A : p_F < p_M$
3. $H_0 : p_F \leq p_M$ and $H_A : p_F > p_M$

The test static is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\hat{\sigma}}$$

where

$$\hat{\sigma} = \sqrt{\frac{\overline{p}(1-\overline{p})}{n_1} + \frac{\overline{p}(1-\overline{p})}{n_2}}$$

The decision regarding rejection and non-rejection of null hypothesis is taken on the basis of significance level α.

Exercise 16: In a study of 29 male and 44 female adults, 11 male and 24 female fell below the characteristic value. Does this study provides sufficient to conclude that females are more likely than males? The significance level is $\alpha = 0.05$.

Solution: For the case stated here, the null hypothesis H_0 and alternate hypothesis H_A are

$$H_0 : p_F \leq p_M \qquad \text{and} \qquad H_A : p_F > p_M$$

Here, we have

$$\hat{p}_F = \frac{24}{44} = 0.545 \qquad \hat{p}_M = \frac{11}{29} = 0.379 \qquad \overline{p} = \frac{24+11}{44+29} = 0.479$$

Now, we have

$$\hat{\sigma} = \sqrt{\frac{\overline{p}(1-\overline{p})}{n_1} + \frac{\overline{p}(1-\overline{p})}{n_2}} = \sqrt{\frac{0.479 \times 0.521}{44} + \frac{0.479 \times 0.521}{29}} = 0.1195$$

This is case of one-tail test on right side. For one-tail test, we have significance level $\alpha = 0.05$. From Table 1 in Appendix, the z-value corresponding to $1 - 0.05 = 0.95$ is 1.645. The rejection region lies between $z = +1.645$ and $z = \infty$. The non-rejection region lies between $z = -\infty$ and $z = +1.645$. The test static is

$$z = \frac{(0.545 - 0.379) - 0}{0.1195} = 1.39$$

Since $z = 1.39$ lies in non-rejection region, the null hypothesis cannot be rejected.

12. Problems and questions

1. What is a hypotheses? Describe characteristics of a hypothesis.
2. Describe the procedure for hypothesis testing.
3. Discuss about the power curve for two-sides test.
4. Discuss about the power curve for one-side test.
5. Write short notes on the following:

    (i) Null hypothesis and alternate hypothesis
    (ii) Significance level
    (iii) Decision rule or test hypothesis
    (iv) Type I and type II errors
    (v) One-tail tests
    (vi) Two-tail test
    (vii) Power curve for one-side test
    (viii) Power curve for two-sides test
    (ix) Sample size to control type II error
    (x) Two sample means
    (xi) Paired comparisons
    (xii) Hypothesis testing for a single population proportion
    (xiii) Hypothesis testing for the difference between two population proportions

VI. Chi-square Tests

Chi-square tests are commonly used in science and technology. In this chapter, we shall discuss about chi-square tests.

1. Chi-square distribution

Chi-square distribution (also called χ^2-distribution) is one of the most widely used probability distributions. The probability density function $f(x,k)$ of chi-square distribution is expressed as

$$f(x,k) = \frac{1}{2^{k/2}\Gamma(k/2)}\ x^{k/2-1}\ \mathrm{e}^{-x/2}$$

where k is the degrees-of-freedom and $x \geq 0$. Variations of $f(x,k)$ as a function of x for six values of k are shown in Figure 1.

2. Chi-square test for comparing variance

Table 3 in Appendices gives percentile of χ^2-distribution for various confidence levels and degrees-of-freedom. These values are often used to decide about rejection or non-rejection of null hypothesis. Suppose a random sample is drawn out of a normal population having variance σ_p^2. For the sample having n items with variance σ_s^2, the test static is

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}\ (n-1)$$

where $(n-1)$ is the degrees-of-freedom and the sample variance is

$$\sigma_s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 \Big/ (n-1)$$

Here, $\overline{x}$ is the arithmetic mean of the values x_i. For the test, the null hypothesis H_0 and alternative hypothesis H_A are

$$H_0 : \sigma_s^2 \leq \sigma_p^2 \qquad \text{and} \qquad H_A : \sigma_s^2 > \sigma_p^2$$

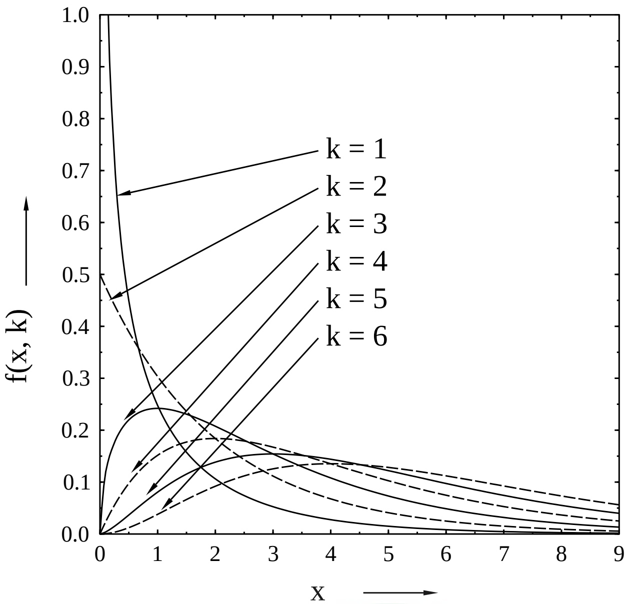


Figure 1: Chi-square distributions for six different values of k.

For the given significance level and degrees-of-freedom, we get critical value from Table 3 in Appendices. When the value of test static is larger than the critical value, the null hypothesis is rejected. On the other side, when the test static is smaller than the critical value, the null hypothesis cannot be rejected.

Exercise 1: Weights kg of 11 students are given in the following table.

S.No.	1	2	3	4	5	6	7	8	9	10	11
Weight	36	42	46	55	45	39	60	49	38	53	35

Make the χ^2-test with 5% and 1% significance levels if we can say that the variance of distribution of weights of all students from which this sample of 11 students is drawn is more than 25 kg?

Solution: For the test the null and alternative hypotheses are

$$H_0 : \sigma_s^2 \leq \sigma_p^2 \qquad \text{and} \qquad H_A : \sigma_s^2 > \sigma_p^2$$

where $\sigma_p^2 = 25$. Let us first calculate the variance σ_s^2 of sample data and for that we prepare the following table.

S. No.	x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$
1	36	−10	100
2	42	−4	16
3	46	0	0
4	58	12	144
5	48	2	4
6	39	−7	49
7	60	14	196
8	49	3	9
9	38	−8	64
10	53	7	49
11	37	−9	81
sum	506		712

The arithmetic mean is

$$\overline{x} = \frac{\sum x_i}{n} = \frac{506}{11} = 46$$

The sample variance is

$$\sigma_s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 \Big/ (n-1) = \frac{712}{11-1} = 71.2$$

The test static is

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}\,(n-1) = \frac{71.2}{25}\,(11-1) = 28.48$$

The degrees-of-freedom is $11 - 1 = 10$. For degrees-of-freedom 10, with 5% significance level, the characteristic value from Table 3 in Appendices is 18.307 and with 1% significance level, the characteristic value is 23.209. Both of these values are smaller than the calculated value of test static. Hence, the null hypothesis is rejected for both significance levels.

Exercise 2: Percentages of marks obtained of 12 students are as given in the following table.

S.No.	1	2	3	4	5	6	7	8	9	10	11	12
Marks	75	81	65	71	54	69	74	59	77	83	62	58

Make the χ^2-test with 5% significance level if we can say that the variance of distribution of percentages of marks of all students from which this sample of 12 students is drawn is more than 80?

Solution: For the test the null and alternative hypotheses are

$$H_0 : \sigma_s^2 \leq \sigma_p^2 \qquad \text{and} \qquad H_A : \sigma_s^2 > \sigma_p^2$$

where $\sigma_p^2 = 80$. Let us first work out the variance σ_s^2 of the sample data with the help of the following table.

S. No.	x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$
1	75	6	36
2	81	12	144
3	65	−4	16
4	71	2	4
5	54	−15	225
6	69	0	0
7	74	5	25
8	59	−10	100
9	77	8	64
10	83	14	196
11	62	−7	49
12	58	−11	121
sum	828		980

The arithmetic average is

$$\overline{x} = \frac{\sum x_i}{n} = \frac{828}{12} = 69$$

The sample variance is

$$\sigma_s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \Big/ (n-1) = \frac{980}{12-1} = 89.09$$

The test static is

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}\,(n-1) = \frac{89.09}{80}\,(12-1) = 12.25$$

The degrees-of-freedom are $12 - 1 = 11$. At 5% significance level, the table gives $\chi^2 = 19.675$. Since the test static is smaller than 19.675, the null hypothesis cannot be rejected.

3. Chi-square test

The chi-square (χ^2) test measures the extent to which the observed frequencies deviate from the corresponding expected frequencies. The test static is

$$\chi^2 = \sum \left[\frac{(f_A - f_e)^2}{f_e}\right]$$

where f_A is the actual frequency, and f_e the expected frequency. For given significance level and degrees-of-level, critical value is obtained from Table 3 in Appendices. Decision for rejection or non-rejection of null hypothesis is taken by comparing the test static with the critical value. When the test static is larger than the critical value, the null hypothesis is rejected. On the other side, when the test static is smaller than the critical value, the null hypothesis cannot be rejected.

Exercise 3 : Following table shows the result of a survey performed in a city in India.

Number of	Hindu	Muslim
families taking tea	1236	164
families not taking tea	564	36

When the expected frequencies are

Number of	Hindu	Muslim
families taking tea	1260	140
families not taking tea	540	60

Considering 5% significance level, decide if there is any significant difference between the two communities in the matter of tea-taking.

Solution: Here, null hypothesis is that there is no difference between the two communities in respect of taking tea. The degrees-of-freedom $= (r-1)(c-1) = (2-1)(2-1) = 1$. At 5% significance level, critical value obtained from Table 3 in Appendices is 3.841. Thus, if the test static is larger than 3.841, the null hypothesis is rejected. We perform χ^2-test and test static is

$$\chi^2 = \sum \left[\frac{(f_A - f_e)^2}{f_e}\right] = \frac{(1236-1260)^2}{1260} + \frac{(164-140)^2}{140}$$

$$+\frac{(564-540)^2}{540} + \frac{(36-60)^2}{60} = 15.22$$

As the test static 15.22 is larger than the critical value 3.841, the null hypothesis is rejected and conclusion is that there is difference between he two communities in respect of taking tea.

Exercise 4: Genetic theory states that children having one parent of blood group A and other of group B will always be of one of three types, A, AB, B and that the proportion of three types on average will be 1 : 2 : 1. A report states that out of 400 children having one parent of blood group A and other of group B, 35% were found of type A, 40 percent of type AB and the remainder of type B. Considering 5% significance level, decide if there is any significant difference from the genetic theory.

Solution: Here, the degrees-of-freedom is $3 - 1 = 2$. With 5% significance level and degrees-of-freedom 2, the critical value obtained form Table 3 in Appendices is 5.991. For actual frequencies (f_A) and expected frequencies (f_e) let us prepare the following table.

Type	f_A	f_e	$(f_A - f_e)$	$(f_A - f_e)^2$	$(f_A - f_e)^2/f_e$
A	140	100	40	1600	16
AB	160	200	-40	1600	8
B	100	100	0	0	0
				sum =	24

The test static is

$$\chi^2 = \sum \left[\frac{(f_A - f_e)^2}{f_e}\right] = 24$$

Since the test static 24 is larger than the critical value 5.991, it does not support the hypothesis of genetic theory that on average types, A, AB, B stand in proportion 1 : 2 : 1.

Exercise 5: Out of 2000 children in a town, 600 children were vaccinated and 1400 children were not vaccinated. Following table shows the number of children attacked and not attacked by smallpox.

	Attacked ($B1$)	Not attacked ($B2$)	Total
Vaccinated ($A1$)	40	560	600
Not vaccinated ($A2$)	190	1210	1400
Total	230	1770	2000

With 5% significance level, apply χ^2 test to decide the effectiveness of vaccination in preventing the attack from smallpox.

Solution: Let us take the null hypothesis H_0 and alternative hypothesis H_A are as the following.

H_0 : Vaccination is effective in preventing the attack from smallpox

H_A : Vaccination is not effective in preventing the attack from smallpox

Total number of children $N = 2000$. Number of children vaccinated $A1$ $= 600$ and not vaccinated $A2 = 1400$; number of children attacked by smallpox $B1 = 230$ and not attacked $B2 = 1770$. Expectation of number of children vaccinated and attacked by smallpox is

$$\text{Expectation of } (A1B1) = \frac{A1 \times B1}{N} = \frac{600 \times 230}{2000} = 69$$

Expectation of number of children vaccinated and not attacked by smallpox is

$$\text{Expectation of } (A1B2) = \frac{A1 \times B2}{N} = \frac{600 \times 1770}{2000} = 531$$

Expectation of number of children not vaccinated and attacked by smallpox is

$$\text{Expectation of } (A2B1) = \frac{A2 \times B1}{N} = \frac{1400 \times 230}{2000} = 161$$

Expectation of number of children not vaccinated and not attacked by smallpox is

$$\text{Expectation of } (A2B2) = \frac{A2 \times B2}{N} = \frac{1400 \times 1770}{2000} = 1239$$

For actual frequencies (f_A) and expected frequencies (f_e) let us prepare the following table.

Group	f_A	f_e	$(f_A - f_e)$	$(f_A - f_e)^2$	$(f_A - f_e)^2/f_e$
A1B1	40	69	−29	841	12.19
A1B2	560	531	29	841	1.58
A2B1	190	161	−29	841	5.22
A2B2	1210	1239	29	841	0.68
				sum =	19.68

The test static is

$$\chi^2 = \sum \left[\frac{(f_A - f_e)^2}{f_e}\right] = 19.68$$

Degrees-of-freedom in this case $= (r-1)(c-1) = (2-1)(2-1) = 1$. For 5% significance level and degrees-of-freedom 1, the critical value obtained from Table 3 in Appendices is 3.841. Since the test static 19.68 is larger than the critical value 3.841, the null hypothesis is rejected. Hence, we conclude that the vaccination is effective in preventing the attack of smallpox.

Exercise 6: One investigator talked 400 people and classified them into three categories, Poor, Middle, Rich. Other investigator talked

600 people and classified them into the same three categories. The information is given in the following table.

Investigator	Poor ($B1$)	Middle ($B2$)	Rich ($B3$)	Total
One ($A1$)	320	60	20	400
Other ($A2$)	280	240	80	600
Total	600	300	100	1000

With 5% significance level, apply χ^2 test to decide that smapling technique of at least one investigator is defective.

Solution: Let us take the null hypothesis H_0 and alternative hypothesis H_A are as the following.

$$H_0 : \text{Technique of at least one investigator is defective}$$

$$H_A : \text{Technique of both investigators is not defective}$$

Total number of children $N = 1000$. Number of people investigated by first investigator $A1 = 400$ and by second investigator $A2 = 600$. Number of poor $B1 = 600$, number of middle $B2 = 300$, and number of rich $B3 = 100$. Expectation of number of poor people for first investigator is

$$\text{Expectation of } (A1B1) = \frac{A1 \times B1}{N} = \frac{400 \times 600}{1000} = 240$$

Expectation of number of middle people for first investigator is

$$\text{Expectation of } (A1B2) = \frac{A1 \times B2}{N} = \frac{400 \times 300}{1000} = 120$$

Expectation of number of rich people for first investigator is

$$\text{Expectation of } (A1B3) = \frac{A1 \times B3}{N} = \frac{400 \times 100}{1000} = 40$$

Expectation of number of poor people for second investigator is

$$\text{Expectation of } (A2B1) = \frac{A2 \times B1}{N} = \frac{600 \times 600}{1000} = 360$$

Expectation of number of middle people for second investigator is

$$\text{Expectation of } (A2B2) = \frac{A2 \times B2}{N} = \frac{600 \times 300}{1000} = 180$$

Expectation of number of rich people for second investigator is

$$\text{Expectation of } (A3B3) = \frac{A3 \times B3}{N} = \frac{600 \times 100}{1000} = 60$$

For actual frequencies (f_A) and expected frequencies (f_e) let us prepare the following table.

Group	f_A	f_e	$(f_A - f_e)$	$(f_A - f_e)^2$	$(f_A - f_e)^2/f_e$
A1B1	320	240	80	6400	26.67
A1B2	60	120	−60	3600	30
A1B3	20	40	−20	400	10
A2B1	280	360	−80	6400	17.78
A2B2	240	180	60	3600	20
A2B3	80	60	20	400	6.67
				sum =	111.12

The test static is

$$\chi^2 = \sum \left[\frac{(f_A - f_e)^2}{f_e}\right] = 111.12$$

Degrees-of-freedom in this case $= (r-1)(c-1) = (2-1)(3-1) = 2$. For 5% significance level and degrees-of-freedom 2, the critical value obtained from Table 3 in Appendices is 5.991. Since the test static 111.12 is larger than the critical value 5.991, the null hypothesis is rejected. Hence, the sampling technique adopted by two investigators differ and that are not similar.

Exercise 7: A dice is thrown 90 times with the following results. For 5% significance level, estimate if the dice is unbiased.

Number turned up	1	2	3	4	5	6
Frequency	18	16	13	17	14	12

Solution: In this case, null hypothesis and alternate hypothesis are

$$H_0 : \text{dice is unbiased} \qquad \text{and} \qquad H_A : \text{dice is biased}$$

Here, the degrees-of-freedom is $6 - 1 = 5$. With 5% significance level and degrees-of-freedom 5, the critical value obtained form Table 3 in Appendices is 11.070. When the dice is unbiased, out of 90 throws, the probability of obtaining any one of the six numbers is 15. For actual frequencies (f_A) and expected frequencies (f_e) let us prepare the following table.

No.	f_A	f_e	$(f_A - f_e)$	$(f_A - f_e)^2$	$(f_A - f_e)^2/f_e$
1	18	15	3	9	0.600
2	16	15	1	1	0.067
3	13	15	-2	4	0.267
4	17	15	2	4	0.267
5	14	15	-1	1	0.067
6	12	15	-3	9	0.600
				sum =	1.867

The test static is

$$\chi^2 = \sum \left[\frac{(f_A - f_e)^2}{f_e}\right] = 1.867$$

Since the test static 1.867 is smaller than the critical value 11.070, the null hypothesis cannot be rejected. We can therefore conclude that the dice may be unbiased.

Exercise 8: A dice is thrown 114 times with the following results. For 5% significance level, estimate if the dice is unbiased.

Number turned up	1	2	3	4	5	6
Frequency	10	21	23	14	29	17

Solution: In this case, null hypothesis and alternate hypothesis are

$$H_0 : \text{dice is unbiased} \qquad \text{and} \qquad H_A : \text{dice is biased}$$

Here, the degrees-of-freedom is $6 - 1 = 5$. With 5% significance level and degrees-of-freedom 5, the critical value obtained form Table 3 in Appendices is 11.070. When the dice is unbiased, out of 114 throws, the probability of obtaining any one of the six numbers is 19. For actual frequencies (f_A) and expected frequencies (f_e) let us prepare the following table.

No.	f_A	f_e	$(f_A - f_e)$	$(f_A - f_e)^2$	$(f_A - f_e)^2/f_e$
1	10	19	-9	81	4.263
2	21	19	2	4	0.211
3	23	19	4	16	0.842
4	14	19	-5	25	1.316
5	29	19	10	100	5.263
6	17	19	-2	4	0.211
				sum =	12.105

The test static is

$$\chi^2 = \sum \left[\frac{(f_A - f_e)^2}{f_e}\right] = 12.105$$

Since the test static 12.105 is larger than the critical value 11.070, the null hypothesis may be rejected. We can therefore conclude that the dice may be biased.

Exercise 9: A packet of 52 playing cards has four types of cards, club ♣, diamond ♢, heart ♡, spade ♠. Individual cards are drawn 104 times and kept back each time in the packet. The results obtained are given in the following table. For 5% significance level, estimate if the process of drawing cards is unbiased.

Type =	club ♣	diamond ♢	heart ♡	spade ♠
Frequency =	20	31	25	28

Solution: In this case, null hypothesis and alternate hypothesis are

H_0 : process is unbiased and H_A : process is biased

Here, the degrees-of-freedom is $4 - 1 = 3$. With 5% significance level and degrees-of-freedom 3, the critical value obtained form Table 3 in Appendices is 7.815. When the process is unbiased, out of 104 draws, the probability of obtaining any one of the four types 26. For actual frequencies (f_A) and expected frequencies (f_e) let us prepare the following table.

Type	f_A	f_e	$(f_A - f_e)$	$(f_A - f_e)^2$	$(f_A - f_e)^2/f_e$
♣	20	26	-6	36	1.385
♢	31	26	5	25	0.962
♡	25	26	-1	1	0.038
♠	28	26	2	4	0.154
				sum =	2.538

The test static is

$$\chi^2 = \sum \left[\frac{(f_A - f_e)^2}{f_e}\right] = 2.538$$

Since the test static 2.538 is smaller than the critical value 7.815, the null hypothesis cannot be rejected. We can therefore conclude that the process may be unbiased.

Exercise 10: A packet of 52 playing cards has thirteen numbers, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K. Individual cards are drawn 104 times and kept back each time in the packet. The results obtained are given in the following table. For 5% significance level, estimate if the process of drawing cards is unbiased.

Number =	1	2	3	4	5	6	7	8	9	10	J	Q	K
Frequency =	4	11	7	10	8	3	9	7	9	13	5	7	11

Solution: In this case, null hypothesis and alternate hypothesis are

H_0 : process is unbiased and H_A : process is biased

Here, the degrees-of-freedom is $13 - 1 = 12$. With 5% significance level and degrees-of-freedom 12, the critical value obtained form Table 3 in Appendices is 21.0026. When the process is unbiased, out of 104 draws, the probability of obtaining any one of the four types 8. For actual frequencies (f_A) and expected frequencies (f_e) let us prepare the following table.

No.	f_A	f_e	$(f_A - f_e)$	$(f_A - f_e)^2$	$(f_A - f_e)^2/f_e$
1	4	8	-4	16	2.000
2	11	8	3	9	1.125
3	7	8	-1	1	0.125
4	10	8	2	4	0.500
5	8	8	0	0	0.000
6	3	8	-5	25	3.125
7	9	8	1	1	0.125
8	7	8	-1	1	0.125
9	9	8	1	1	0.125
10	13	8	5	25	3.125
11	5	8	-3	9	1.125
12	7	8	-1	1	0.125
13	11	8	3	9	1.125
				sum =	12.75

The test static is

$$\chi^2 = \sum \Big[\frac{(f_A - f_e)^2}{f_e}\Big] = 12.75$$

Since the test static 12.75 is smaller than the critical value 21.026, the null hypothesis cannot be rejected. We can therefore conclude that the process may be unbiased.

4. Problems and questions

1. Monthly salaries of 10 workers in rupees are given in the following table.

S.No.	1	2	3	4	5	6	7	8	9	10
Salary	360	425	465	550	455	395	620	490	380	530

Make the χ^2-test with 5% significance level if we can say that the variance of distribution of monthly salaries of all workers from which this sample of 10 workers is drawn is more than 250 rupees?

2. Genetic theory states that children having one parent of blood group A and other of group B will always be of one of three types, A, AB, B and that the proportion of three types on average will be 1 : 2 : 1. A report states that out of 200 children having one parent of blood group A and other of group B, 30% were found of type A, 45 percent of type AB and the remainder of type B. Considering 5% significance level, decide if there is any significant difference from the genetic theory.

3. What is Chi-square test? A dice is thrown 102 times with the following results. For 5% significance level, estimate if the dice is unbiased.

Number turned up	1	2	3	4	5	6
Frequency	20	18	15	19	16	14

4. A packet of 52 playing cards has four types of cards, club ♣, diamond ♢, heart ♡, spade ♠. Individual cards are drawn 156 times and kept back each time in the packet. The results obtained are given in the following table. For 5% significance level, estimate if the process of drawing cards is unbiased.

5. Write short notes on the following:

   (i) Chi-square test for comparing variance

   (ii) Chi-square distribution

   (iii) Chi-square test

VII. Non-parametric Tests

As mentioned earlier, statistical tests may be classified into two categories: (i) parametric tests and (ii) non-parametric tests. The first meaning of non-parametric covers techniques that do not rely on data belonging to any particular distribution. The second meaning of non-parametric covers techniques that do not assume that the structure of a model is fixed. Non-parametric models differ from parametric models in that the model structure is not specified a priori but is instead determined from data. The term non-parametric does not mean to imply that such models completely lack parameters but that the number and nature of the parameters are flexible and not fixed in advance. Some of the non-parametric tests are as the following.

(i) Sign test

(ii) Run test

(iii) Wilcoxon test

(iv) Median test

(v) Mann-Whitney U test

(vi) Kolmogorov-Smirnov test

In the present chapter, we shall discuss about these tests.

1. Sign test

The sign-test is one of non-parametric tests. Its name comes from the fact that it is based on plus and minus signs, rather than the values of observations in the sample. The sign-test may be applied for the following two situations:

(i) For a single sample

(ii) For a paired sample

We shall discuss about them separately.

1.1 Single sample sign test

In the sign test, for some hypothesized population median μ_0, we assign a plus sign (+) when the given value is larger than μ_0 and a minus sign (−) when the given value is smaller than μ_0. When the given value is equal to μ_0, we ignore that value and do not account for in the test. For the sign test, the possible null and alternative hypotheses may be as the following.

(a) $H_0 : P(+) = P(-)$ and $H_A : P(+) \neq P(-)$

(b) $H_0 : P(+) \geq P(-)$ and $H_A : P(+) < P(-)$

(c) $H_0 : P(+) \leq P(-)$ and $H_A : P(+) > P(-)$

where $P(+)$ and $P(-)$ denote the probabilities of getting plus signs and minus signs, respectively. For two-sides alternative hypothesis $H_A : P(+) \neq P(-)$, the test statistic is the smaller of the number of plus and minus signs. For one-side alternative hypothesis $H_A : P(+) > P(-)$, the test statistic is the number of minus signs. For the other one-side alternative hypothesis $H_A : P(+) < P(-)$, the test statistic is the number of plus signs. This number of signs is denoted by x.

The test static is the probability of observing x signs when a given sample of effective size n and parameter p is evaluated by the expression

$$P = \sum_{k=0}^{x} {}^nC_k \; p^k \; q^{n-k}$$

where $q = 1-p$. For the significance level α, we reject the null hypothesis when P is less than $\alpha/2$ (in case of two-sides test) or less than α (in case of one-side test). Further, the p-value[1] is $2P$ in case of two-sides test and is P in case of one-side test.

Exercise 1: For randomly selected sample of 10 girls, the scores obtained are given in the following table. Apply the sign-test with 5% significance level to decide if the median score of the population from which this sample is drawn is different from 5. Consider equal probability for plus and minus signs.

[1] Distinguish between the letters p used for the parameter p and and for the p-value.

Girl	Score	Girl	Score
1	4	6	6
2	5	7	10
3	8	8	7
4	8	9	6
5	9	10	6

Solution: The null and alternative hypotheses are

$$H_0 : \mu = 5 \qquad \text{and} \qquad H_A : \mu \neq 5$$

They are equivalent to

$$H_0 : P(+) = P(-) \qquad \text{and} \qquad H_A : P(+) \neq P(-)$$

We first examine the given data to determine which scores lie above and which ones lie below the hypothesized median of 5. When we assign a plus sign to those scores which lie above the hypothesized median and a minus sign to those which lie below the hypothesized median, the results are shown in the following table. For the scores which are equal to the hypothesized median, no plus or minus sign is assigned, and those values are not accounted for in the test.

Girl	Score	Girl	Score
1	−	6	+
2	0	7	+
3	+	8	+
4	+	9	+
5	+	10	+

In fact, we expect the number of scores falling above and below 5 to be approximately equal. Thus, the probability of plus is equal to the probability of minus, and both probabilities are equal to 0.5. That is, $p = q = 0.5$. Here, the number of plus signs is 8 and minus signs is 1. Therefore, $n = 8 + 1 = 9$ and $x = 1$. The test static is

$$P = \sum_{k=0}^{1} {}^{9}C_k \; p^k \; q^{n-k} = \; {}^{9}C_0 \; (0.5)^0 \; (0.5)^9 + \; {}^{9}C_1 (0.5)^1 \; (0.5)^8$$

$$= 0.00195 + 0.01758 = 0.0195$$

For two-sides significance level with 5%, we have $\alpha = 0.05$ and $\alpha/2 = 0.025$. Since the test static 0.0195 is less than the critical value 0.025, the null hypothesis is rejected and therefore, we conclude that the median score is not 5. The p value for this test is 2 (0.0195) = 0.039.

Exercise 2: For a randomly selected sample of 12 students, the scores obtained are as given in the following table. Apply the sign-test with 6% significance level to decide if the median score of the population from which this sample is drawn is more than 10. Consider equal probability for plus and minus signs.

Student	Score	Student	Score
1	10	7	12
2	12	8	13
3	8	9	7
4	11	10	14
5	9	11	10
6	11	12	14

Solution: The null and alternative hypotheses are

$$H_0 : \mu \leq 10 \qquad \text{and} \qquad H_A : \mu > 10$$

They are equivalent to

$$H_0 : P(+) \leq P(-) \qquad \text{and} \qquad H_A : P(+) > P(-)$$

We first examine the given data to determine which scores lie above and which ones lie below the hypothesized median of 10. When we assign a plus sign to those scores which lie above the hypothesized median and a minus sign to those which lie below the hypothesized median, the results are shown in the following table. For the scores which are equal to the hypothesized median, no plus or minus sign is assigned, and those values are not accounted for in the test.

Student	Score	Student	Score
1	0	7	+
2	+	8	+
3	−	9	−
4	+	10	+
5	−	11	0
6	+	12	+

In fact, we expect the number of scores falling above and below 10 to be approximately equal. Thus, the probability of plus is equal to the probability of minus, and those probabilities are equal to 0.5. That is, $p = q = 0.5$. Here,the number of plus signs is 7 and minus signs is 3. Therefore, $n = 7 + 3 = 10$ and $x = 3$. The test static is

$$P = \sum_{k=0}^{3} {}^{10}C_k \; p^k \; q^{n-k} = \; {}^{10}C_0 \; (0.5)^0 (0.5)^{10} + {}^{10}C_1 \; (0.5)^1 (0.5)^9$$

$$+^{10}C_2\ (0.5)^2(0.5)^8 +^{10}C_3\ (0.5)^3(0.5)^7 = 9.7656 \times 10^{-4}$$

$$+9.7656 \times 10^{-3} + 4.3945 \times 10^{-2} + 0.11718 = 0.1719$$

For one-side significance level with 6%, we have $\alpha = 0.06$. Since the test static 0.1719 is less than critical value 0.06, the null hypothesis is rejected and therefore, we conclude that the median score may be more than 10. The p value for this test is 0.1719.

1.2 Sign test for paired data

Here, we consider two sets of data. One of the matched score, say Y_i, is subtracted from the other score, say X_i. When Y_i is less than X_i, the sign of difference is positive, and when Y_i is more than X_i, the sign of difference is negative. When the median difference is zero, we do not consider that value in the test. After getting plus and minus signs, the remaining procedure is the same as discussed for the single sample sign test.

Exercise 3: For the data given in the following table, perform the sign-test for the significance level 5% and the hypotheses

$$H_0 : P(+) \geq P(-) \qquad \text{and} \qquad H_A : P(+) < P(-)$$

$X_i =$	4.2	5.6	8.3	8.7	9.2	6.5	10.0	7.3	6.6	6.4
$Y_i =$	6.5	5.6	9.2	10.5	5.6	12.3	13.7	8.4	9.5	6.4

Solution: We prepare the following table where plus sign is given when $X_i > Y_i$ and minus sign is given when $X_i < Y_i$. When $X_i = Y_i$, we have written 0 and will ignore this entry in the test. There are one plus and seven minus signs. Thus, $n = 1 + 7 = 8$. For the given hypotheses, we decide on the basis of the number of plus signs. Thus, $x = 1$. Considering equal probabilities for plus and minus signs, we have $p = q = 0.5$. The test static is

$$P = \sum_{k=0}^{1} {}^8C_k\ p^k\ q^{n-k} = {}^8C_0\ (0.5)^0\ (0.5)^8 +^8C_1(0.5)^1\ (0.5)^7$$

$$= 0.0039 + 0.0312 = 0.0351$$

Sr. No.	Score X_i	Y_i	Sign
1	4.2	6.5	–
2	5.6	5.6	0
3	8.3	9.2	–
4	8.7	10.5	–
5	9.2	5.6	+
6	6.5	12.3	–
7	10.0	13.7	–
8	7.3	8.4	–
9	6.6	9.5	–
10	6.4	6.4	0

With $\alpha = 0.05$, for one-side test, we have critical value 0.05. Since the test static 0.0351 is smaller than the critical value 0.05, the null hypothesis is rejected. The p-value is 0.0351.

Exercise 4: For the data given in the following table, perform the sign-test for the significance level 5% and the hypotheses

$$H_0 : P(+) \leq P(-) \qquad \text{and} \qquad H_A : P(+) > P(-)$$

$X_i =$	13.5	17.8	18.3	35.2	20.8	38.6	19.2	19.8	38.0	59.8
$Y_i =$	14.7	16.6	18.3	28.8	22.2	49.3	16.7	18.5	37.3	46.8

Solution: We prepare the following table where plus sign is given when $X_i > Y_i$ and minus sign is given when $X_i < Y_i$. When $X_i = Y_i$, we have written 0 and would ignore this entry in the test. There are six plus and three minus signs. Thus, $n = 3 + 6 = 9$. For the given hypotheses, we decide on the basis of the number of minus signs. Thus, $x = 3$.

Sr. No.	Score X_i	Y_i	Sign
1	13.5	14.7	–
2	17.8	16.6	+
3	18.3	18.3	0
4	35.2	28.8	+
5	20.8	22.2	–
6	38.6	49.3	–
7	19.2	16.7	+
8	19.8	18.5	+
9	38.0	37.3	+
10	59.8	46.8	+

Considering equal probabilities for plus and minus signs, we have $p = q = 0.5$. The test static is

$$P = \sum_{k=0}^{3} {}^9C_k \; p^k \; q^{n-k} = \; {}^9C_0 \; (0.5)^0(0.5)^9 + {}^9C_1 \; (0.5)^1(0.5)^8 + {}^9C_2 \; (0.5)^2(0.5)^7$$

$$+ {}^9C_3 \; (0.5)^3(0.5)^6 = 0.0020 + 0.0176 + 0.0703 + 0.1641 = 0.2539$$

With $\alpha = 0.05$, for one-side test, we have critical value 0.05. Since test static 0.2539 is larger than critical value 0.05, the null hypothesis cannot be rejected. The p-value is 0.2539.

Exercise 5: For the data given in the following table, perform the sign-test for the significance level 5% and the hypotheses

$$H_0 : P(+) = P(-) \qquad \text{and} \qquad H_A : P(+) \neq P(-)$$

$X_i =$	14.7	17.9	17.2	25.8	23.9	39.5	17.7	10.4	38.6	44.7
$Y_i =$	14.9	16.3	19.3	24.9	25.2	39.5	18.6	15.7	38.4	45.7

Solution: We prepare the following table where plus sign is given when $X_i > Y_i$ and minus sign is given when $X_i < Y_i$. When $X_i = Y_i$, we have written 0 and would ignore this entry in the test. There are three plus and six minus signs. Thus, $n = 3 + 6 = 9$. For the given hypotheses, we decide on the basis of the minimum of the number of minus and the number of plus signs. Thus, $x = 3$.

Sr.	Score		Sign
No.	X_i	Y_i	
1	14.7	14.9	$-$
2	17.9	16.3	$+$
3	17.2	19.3	$-$
4	25.8	24.9	$+$
5	23.9	25.2	$-$
6	39.5	39.5	0
7	17.7	18.6	$-$
8	10.4	15.7	$-$
9	38.6	38.4	$+$
10	44.7	45.7	$-$

Considering equal probabilities for plus and minus signs, we have $p = q = 0.5$. The test static is

$$P = \sum_{k=0}^{3} {}^9C_k \; p^k \; q^{n-k} = \; {}^9C_0 \; (0.5)^0(0.5)^9 + {}^9C_1 \; (0.5)^1(0.5)^8 + {}^9C_2 \; (0.5)^2(0.5)^7$$

$$+ \, ^9C_3 \, (0.5)^3 (0.5)^6 = 0.0020 + 0.0176 + 0.0703 + 0.1641 = 0.2539$$

With $\alpha = 0.05$, for two-sides test, we have critical value 0.025. Since the test static 0.2539 is larger than the critical value 0.025, the null hypothesis cannot be rejected. The p-value is $2\ P = 0.5078$.

2. Wilcoxon signed rank test

In this test, all possible null and alternative hypotheses, which may be tested about some unknown population mean μ_0, are as the following.

(a) $H_0 : \mu = \mu_0$ and $H_A : \mu \neq \mu_0$

(b) $H_0 : \mu \geq \mu_0$ and $H_A : \mu < \mu_0$

(c) $H_0 : \mu \leq \mu_0$ and $H_A : \mu > \mu_0$

For the given n random observations, the Wilcoxon procedure is as the following.

(i) We subtract the hypothesized mean μ_0 from each of the observation x_i and obtain

$$d_i = x_i - \mu_0$$

When $d_i = 0$ (*i.e.*, $x_i = \mu_0$), eliminate that d_i (*i.e.*, x_i) from the calculations, reduce the number of data n accordingly.

(ii) We obtain $|d_i|$ and now rank them in a sequence from the smallest to the largest values of $|d_i|$. If two or more of the $|d_i|$ are equal, we assign each of the tied values the mean of the rank positions occupied by the tied values. If, for example, the three smallest $|d_i|$ are all equal, place them in rank positions 1, 2, 3, but assign each a rank of $(1+2+3)/3 = 2$.

(iii) Assign the sign of d_i to the corresponding rank to get the signed rank.

(iv) Find T_+, the sum of the ranks with the positive sign and T_-, the sum of the ranks with the negative sign. We cannot expect T_+ and T_- computed from a given sample to be equal.

(v) The test statistic is either T_+ or T_-, depending on the nature of the alternative hypothesis.

When the alternative hypothesis is two sides ($\mu \neq \mu_0$), either a sufficiently small value of T_+ or a sufficiently small value of T_- will cause us to reject the null hypothesis H_0 and smaller of the T_+ and T_- is the test statistic.

When H_A: $\mu < \mu_0$, we expect our sample to yield a large value of T_+. Therefore, when the one-side alternative hypothesis states that the true population mean is less than the hypothesized mean, a sufficiently small value of T_+ will cause rejection of the null hypothesis H_0 and T_+ is the test statistic.

When H_A: $\mu > \mu_0$, we expect our sample to yield a large value of T_-. Therefore, when the one-side alternative hypothesis states that the true population mean is more than the hypothesized mean, a sufficiently small value of T_- will cause rejection of H_0 and T_- is the test statistic.

(vi) Corresponding to the significance level α and effective number n of data, we obtained critical value from Table 5 in Appendices. When the test static is larger than the critical value, the null hypothesis cannot be rejected. On the other side, when the test static is smaller than the critical value, the null hypothesis is rejected.

Exercise 6: For a random sample of 15 patients, the results are as the following.

4.91	4.10	6.74	7.27	7.42	7.50	6.56	4.64
5.98	3.14	3.23	5.80	6.17	5.39	5.77	

We wish to know if we can conclude on the basis of these data that the population mean is different from 5.05.

Solution: The hypotheses are

$$H_0 : \mu = 5.05 \qquad \text{and} \qquad H_A : \mu \neq 5.05$$

(i) We prepare a table where in the first column, we write down data x_i. In second column we calculate $d_i = x_i - \mu_0 = x_i - 5.05$ and in third column we, report $|d_i|$. In fourth column, we calculate the rank of $|d_i|$. Finally, the fifth column is the signed rank.

Calculation of test statistic

x_i	d_i	$\|d_i\|$	Rank of $\|d_i\|$	Signed rank
4.91	−0.14	0.14	1	−1
4.10	−0.95	0.95	7	+7
6.74	+1.69	1.69	10	+10
7.27	+2.22	2.22	13	+13
7.42	+2.37	2.37	14	+14
7.50	+2,45	2,45	15	+15
6.56	+1.51	1.51	9	+9
4.64	−0.41	0.41	3	−3
5.98	+0.93	0.93	6	+6
3.14	−1.91	1.91	12	−12
3.23	−1.82	1.82	11	−11
5.80	+0.75	0.75	5	+5
6.17	+1.12	1.12	8	+8
5.39	+0.34	0.34	2	+2
5.77	+0.72	0.72	4	+4

(ii) Sum of positive ranks and negative ranks are, respectively, $T_+ = 86$ and $T_- = 34$. We take the test statistic T as the smaller of T_+ and T_-. Thus, the test statistic $T = 34$.

(iii) With 5% significance level, we have $\alpha = 0.05$. For two-sides alternative hypothesis, we have $\alpha/2 = 0.025$. In Table 5 in Appendices, for $n = 15$, a value close to 0.025 is 0.0240 for which T is 25. Thus, the critical value is 25.

(iv) Since the test static $T = 34$ is larger than the critical value $T = 25$, we cannot reject H_0. We conclude that the population mean may be 5.05.

(v) For $T = 34$ and $n = 15$, Table 5 in Appendices gives 0.0757. Thus, the p value is $p = 2 \times 0.0757 = 0.1514$.

Exercise 7: Using the significance level 0.06, for the following data, perform the Wilcoxon test to conclude if the population mean is larger than 8.05.

9.15 6.21 7.23 8.05 8.39 6.65 7.08 8.05 6.21 8.45

7.01 8.55 7.71 9.23 4.35 6.74 8.71 8.39 7.06

Solution: The hypotheses are

$$H_0 : \mu \leq 8.05 \qquad \text{and} \qquad H_A : \mu > 8.05$$

Now, we prepare the following table where $d_i = x_i - \mu_0 = x_i - 8.05$.

x_i	d_i	$\lvert d_i \rvert$	Rank	Signed Rank
9.15	1.10	1.10	11	11
6.21	−1.84	1.84	15.5	−15.5
7.23	−0.82	0.82	7	−7
8.05	0.00	0.00		
8.39	0.34	0.34	2	2
6.65	−1.40	1.40	14	−14
7.08	−0.97	0.97	8	−8
8.05	0.00	0.00		
6.21	−1.84	1.84	15.5	−15.5
8.45	0.40	0.40	4	4
7.01	−1.04	1.04	10	−10
8.55	0.50	0.50	5	5
7.71	−0.34	0.34	2	−2
9.23	1.18	1.18	12	12
4.35	−3.70	3.70	17	−17
6.74	−1.31	1.31	13	−13
8.71	0.66	0.66	6	6
8.39	0.34	0.34	2	2
7.06	−0.99	0.99	9	−9

Since two d_i's are zero, therefore number of items $n = 19 - 2 = 17$. According to the hypotheses, it is a one-side case, we shall decided on the basis of T_-. For $n = 17$ with significance level 0.06, we have $T = 43$. From the table the sum of positive-rank values is $T_+ = 41$ and the sum of negative-rank values is $T_- = 111$. Since $43 < 111$, the null hypothesis cannot be rejected.

Exercise 8: Using the significance level 0.06, for the following data, perform the Wilcoxon test to conclude if the population mean is smaller than 7.45.

9.13	6.31	8.23	7.45	7.69	6.65	8.08	8.05	7.21	8.75
8.01	8.55	7.71	7.33	4.35	6.74	8.71	7.45	7.06	

Solution: The hypotheses are

$$H_0 : \mu \geq 7.45 \qquad \text{and} \qquad H_A : \mu < 7.45$$

Now, we prepare the following table where $d_i = x_i - \mu_0 = x_i - 7.45$.

x_i	d_i	$\|d_i\|$	Rank	Signed Rank
9.13	1.68	1.68	16	16
6.31	−1.14	1.14	13	−13
8.23	0.78	0.78	10	10
7.45	0.00	0.00		
7.69	0.24	0.24	2.5	2.5
6.65	−0.80	0.80	11	−11
8.08	0.63	0.63	8	8
8.05	0.60	0.60	7	7
7.21	−0.24	0.24	2.5	−2.5
8.75	1.30	1.30	15	15
8.01	0.56	0.56	6	6
8.55	1.10	1.10	12	12
7.71	0.26	0.26	4	4
7.33	−0.12	0.12	1	−1
4.35	−3.10	3.10	17	−17
6.74	−0.71	0.71	9	−9
8.71	1.26	1.26	14	14
7.45	0.00	0.00		
7.06	−0.39	0.39	5	−5

Since two d_i's are zero, therefore number of items $n = 19 - 2 = 17$. According to the hypotheses, it is a one-side case, we shall decided on the basis of T_+. For $n = 17$ with significance level 0.06, we have $T = 43$. From the table the sum of positive-rank values is $T_+ = 94.5$ and the sum of negative-rank values is $T_- = 58.5$. Since $43 < 94.5$, the null hypothesis cannot be rejected.

3. Median test

This is a nonparametric test that may be used to test if the medians of populations from which two samples are drawn are identical. The data in each sample are assigned to two groups, one consisting of data whose values are higher than the median value in the two groups combined, and the other consisting of data whose values are at the median or below. From given two samples, we prepare a table as the following.

	Sample 1	Sample 2	Total
No. of scores above median	a	b	$a+b$
No. of scores on or below median	c	d	$c+d$
Total	$a+c$	$b+d$	n

where $n = a + b + c + d$. The test static is[2]

$$\chi^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}$$

The null and alternate hypotheses are

$$H_0 : M_U = M_R \qquad \text{and} \qquad H_A : M_U \neq M_R$$

For the significance level α, the critical value is obtained from χ^2-table corresponding to degrees-of-freedom 1. When the test static is less than the critical value, the null hypothesis cannot be rejected. On the other side, when the test static is more than the critical value, the null hypothesis is rejected.

Exercise 9: A random sample of 12 male students from a rural junior high school and an independent random sample of 16 male students from an urban junior high school were given a test to measure their level of mental health. The result are given in the following table.

Urban	Rural	Urban	Rural
35	29	25	50
26	50	27	37
27	43	45	34
21	22	46	31
27	42	33	
38	47	26	
23	42	46	
25	32	41	

[2] When the result is in the form of a contingency table consisting of two rows and two columns, such a table is commonly referred to as 2×2 table, expressed as the following. Here, a, b, c, d are the observed cell frequencies.

A 2×2 contingency table

Second criterion of classification	First criterion of classification: 1	2	Total
1	a	b	$a + b$
2	c	d	$c + d$
Total	$a + c$	$b + d$	n

The value of test static χ^2 may be calculated as

$$\chi^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}$$

When we apply the $(r-1)(c-1)$ rule for finding out the degrees-of-freedom to a 2×2 table, the result is 1.

Use median test with 5% significance level to decide if urban and rural male junior high school students differ with respect to their level of mental health.

Solution: To determine if there is difference between two samples, we perform a median test. For median score M_U of sampled population of urban students, and median score M_R of sampled population of rural students, the null and alternative hypotheses are

$$H_0 : M_U = M_R \qquad \text{and} \qquad H_A : M_U \neq M_R$$

For the significance level $\alpha = 0.05$ and degrees-of-freedom 1, critical value is 3.841. Thus, we shall reject H_0 when the test static is larger than the critical value 3.841. After arranging all the 28 (=16+12) data in the increasing sequence, the median is obtained as $(33 + 34)/2 = 33.5$. Now, we can prepare the following table.

	Urban	Rural	Total
No. of scores above median	6	8	14
No. of scores on or below median	10	4	14
Total	16	12	28

The test static is

$$\chi^2 = \frac{28[(6)(4) - (8)(10)]^2}{(16)(12)(14)(14)} = 2.33$$

Since the test static 2.33 is smaller than the critical value 3.841, we cannot reject the null hypothesis. We conclude that the two samples might have been drawn from populations with equal median.

Exercise 10: A random sample of daily income of 16 workers from a factory and an independent random sample of daily income of 14 workers from another factory are given in the following table.

Urban	Rural	Urban	Rural
36	28	30	29
28	48	28	36
45	41	44	42
30	33	46	35
34	45	47	28
28	38	51	32
32	45	29	
49	36	32	

Use median test with 5% significance level to decide if two samples differ with respect to daily income.

Solution: To determine if there is difference between two samples, we perform a median test. For median score M_U of sampled population of urban students, and median score M_R of sampled population of rural students, the null and alternative hypotheses are

$$H_0 : M_U = M_R \qquad \text{and} \qquad H_A : M_U \neq M_R$$

For the significance level $\alpha = 0.05$, critical value of χ^2 with degrees-of-freedom 1 is 3.841. Thus, we shall reject H_0 if the computed χ^2 is ≥ 3.841. After arranging all the 30 (=16+14) data in the increasing sequence, the median is obtained as $(35 + 36)/2 = 35.5$. Now, we can prepare the following table.

	Urban	Rural	Total
No of scores above median	7	8	15
No of scores on or below median	9	6	15
Total	16	14	30

The test static is

$$\chi^2 = \frac{30[(7)(6) - (8)(9)]^2}{(16)(14)(15)(15)} = 0.54$$

Since the test static 0.54 is smaller than the critical value 3.841, we cannot reject the null hypothesis. We conclude that the two samples might have been drawn from populations with equal median.

4. Mann-Whitney U test

The median test does not make full use of all information present in the two samples when the variable of interest is measured. Reducing an observation's information content to merely that of whether or not it falls above or below the common median is a waste of information. When, for testing the desired hypothesis, there is available a procedure that makes use of more of information inherent in the data, that procedure should be used if possible. Such a non-parametric procedure that can often be used instead of median test is known as Mann-Whitney U test. It is sometimes called as Mann-Whitney-Wilcoxon (MWW) test or Wilcoxon rank-sum test, as it based on the ranks of the observations; it uses more information than does the median test.

In the Mann-Whitney U test, the possible null and alternative hypotheses which may be tested about the median M_X of X population and the median M_Y of Y population are as the following.

(a) $H_0 : M_X = M_Y$ and $H_A : M_X \neq M_Y$

(b) $H_0 : M_X \leq M_Y$ and $H_A : M_X > M_Y$

(c) $H_0 : M_X \geq M_Y$ and $H_A : M_X < M_Y$

For the significance level α, the rejection regions are decided as the following. For two-sides alternative hypothesis ($H_A : M_X \neq M_Y$), the rejection regions are from $-\infty$ to $w_{\alpha/2}$ and from $w_{1-J\alpha/2}$ to ∞. For one-side alternative hypothesis ($H_A : M_X > M_Y$), the rejection region is from $w_{1-\alpha}$ to ∞ and for the alternative hypothesis ($H_A : M_X < M_Y$), the rejection region is from $-\infty$ to w_α.

To compute test statistic, we combine the two samples and rank all observations X and Y jointly from the smallest to the largest while keeping the ranks of X and Y separately. The tied observations are assigned a rank equal to the mean of the rank positions for which they are tied. The test statistic is

$$T = S - \frac{n(n+1)}{2} \tag{7.1}$$

where n is the number of sample X observations and S is the sum of the ranks assigned to the sample observations from the population of X values. The choice of which sample's values we label X is arbitrary.

Exercise 11: Let us design an experiment in order to assess the effects of prolonged inhalation of cadmium oxide. Fifteen laboratory animals served an experimental subjects, while ten similar animals served as controls. The variable of interest was hemoglobin level following the experiment. The results are shown in the following table. With the 5% significance level, we wish to know if we can conclude that median of X is smaller than median of Y.

Hemoglobin determinations (gm) for 25 laboratory animals

Exposed (X)	Unexposed (Y)	Exposed (X)	Unexposed (Y)
14.4	17.4	14.1	16.3
14.2	16.2	15.3	16.8
13.8	17.1	15.7	
16.5	17.5	16.7	
14.1	15.0	13.7	
16.6	16.0	15.3	
15.9	16.9	14.0	
15.6	15.0		

Solution: The null and alternative hypotheses are

$$H_0 : M_X \geq M_Y \qquad \text{and} \qquad H_A : M_X < M_Y$$

where M_X is median of population of animals exposed to cadmium oxide and M_Y is median of population of animals not exposed to cadmium oxide. For $n = 15$ and $m = 10$ with one-side hypothesis ($H_A : M_X < M_Y$) and $\alpha = 0.05$, the rejection region is form $-\infty$ to $T = 45$ (Table 6 in Appendices).

To compute test statistic, we combine the two samples and rank all observations X and Y jointly from the smallest to the largest while keeping the ranks of X and Y separately. The tied observations are assigned a rank equal to the mean of the rank positions for which they are tied. The results of this step are shown in that following table.

Given data and ranks

X	Rank	Y	Rank	X	Rank	Y	Rank
13.7	1			15.9	14		
13.8	2					16.0	15
14.0	3					16.2	16
14.1	4.5					16.3	17
14.1	4.5			16.5	18		
14.2	6			16.6	19		
14.4	7			16.7	20		
		15.0	8.5			16.8	21
		15.0	8.5			16.9	22
15.3	10.5					17.1	23
15.3	10.5					17.4	24
15.6	12					17.5	25
15.7	13						
Total					145		180

The test statistic is

$$T = S - \frac{n(n+1)}{2} = 145 - \frac{15(15+1)}{2} = 25$$

Since the calculated value $T = 25$ lies in the rejection region, we reject the null hypothesis H_0. We conclude that M_X is smaller than M_Y.

4.1 Large-sample approximation

When either n or m is so large that we cannot use Table 6 in Appendices for getting the critical values for Mann-Whitney U test. For such case, we compute

$$z = \frac{T - mn/2}{\sqrt{nm(n+m+1)/12}}$$

and compare the result, for significance, with critical values of the standard normal distribution. Here, T can be calculated with the help of equation (7.1).

5. Kolmogorov-Smirnov test

It is also known as the 'goodness-of-fit test'. When we wish to know how well the distribution of sample data conform to some theoretical distribution, a test known as the Kolmogorov-Smirnov goodness-of-fit test provides an alternative to the chi-square goodness-of-fit test. Kolmogorov's work is concerned with the one-sample case whereas Smirnov's work deals with the case involving two samples in which interest centers on testing the hypothesis that the distribution of the two-parent populations are identical. The test for the first situation is often known as the Kolmogorov-Smirnov one-sample test. The test for the two-sample case, is known as the Kolmogorov-Smirnov two-sample test.

In using the Kolmogorov-Smirnov goodness-of-fit test, a comparison is made between some theoretical cumulative distribution function $F_T(x)$ and a sample cumulative distribution function $F_S(x)$. The sample is a random sample from a population with unknown cumulative distribution function $F(x)$.

The difference between theoretical cumulative distribution function $F_T(x)$ and sample cumulative distribution function $F_S(x)$ is measured by the statistic D, which is the greatest vertical distance between $F_S(x)$ and $F_T(x)$. When a two-sides test is appropriate, that is, when null hypothesis H_0 and alternative hypothesis H_A are

$$H_0 : F(x) = F_T(x) \quad \text{for all } x \text{ from } -\infty \text{ to } \infty$$

$$H_A : F(x) \neq F_T(x) \quad \text{for at least one } x$$

The test statistic is

$$D = \overset{sup}{x} \; |F_S(x) - F_T(x)|$$

which is read, "D equals the supremum (greatest), over all x, of the absolute value of the difference $F_S(x)$ minus $F_T(x)$." For significance level α and sample size n, critical value is obtained from Table 8 in Appendices for $1 - \alpha$ (two-sides).

Exercise 12: For 5% significance level, perform Kolmogorov-Smirnov test for the data given in the following table, in order to know if we may conclude that these data are not from a normally distributed population with arithmetic mean 60 and standard deviation 6.

55	64	60	57	48	67	72	57	72	66	58	56
60	61	52	57	72	60	61	57	57	72	48	67
64	55	58	60	60	57	52	61	56	58	61	66

Solution: There are 36 data and therefore $n = 36$. For two-sides hypotheses, significance level $\alpha = 0.05$ and $n = 36$, the critical value from Table 8 in Appendices is 0.221. We shall reject the null hypothesis H_0 when the test static D exceeds the critical value 0.221. The first step for calculation is to prepare the following table.

x	Freq	C. F.	F_S	z	F_T	$\|F_S - F_T\|$
48	2	2	0.0556	−2.00	0.0228	0.0328
52	2	4	0.1111	−1.33	0.0918	0.0193
55	2	6	0.1667	−0.83	0.2033	0.0366
56	2	8	0.2222	−0.67	0.2514	0.0292
57	6	14	0.3889	−0.50	0.3085	0.0804
58	3	17	0.4722	−0.33	0.3707	0.1015
60	5	22	0.6111	0.00	0.5000	0.1111
61	4	26	0.7222	0.17	0.5675	0.1547
64	2	28	0.7778	0.67	0.7486	0.0292
66	2	30	0.8333	1.00	0.8413	0.0080
67	2	32	0.8889	1.17	0.8790	0.0099
72	4	36	1.0000	2.00	0.9772	0.0228
	36					

The first column has data, second column has frequency of data and third column has cumulative frequency. Now, $F_S(x)$ is obtained by dividing cumulative frequencies by the sample size, 36. For arithmetic mean 60 and standard deviation 6, we convert each observed value of x to a value of standard normal variable z through the relation $z = (x-60)/6$, given in fifth column. Form Table 1 in Appendices, we then obtain values of normal distribution integral for the limits from $-\infty$ to z. This value is $F_T(x)$, given in sixth column of the table.

The test statistic D can be computed algebraically or graphically. The possible values of $|F_S - F_T|$ are given in the last column in the table. This table shows that the value of test static D is 0.1547. Since the test static 0.1547 is less than the critical value 0.221, the null hypothesis cannot be rejected. Hence the sample may have come from the specified distribution.

Exercise 13: For 5% significance level, perform Kolmogorov-Smirnov test for the data given in the following table, in order to know if we may conclude that these data are not from a normally distributed population with arithmetic mean 85 and standard deviation 7.

86	95	91	88	79	98	75	88	75	97	89	87
91	92	83	88	70	91	91	89	88	70	79	98
83	86	89	91	91	88	83	92	87	89	92	97

Solution: There are 36 data and therefore $n = 36$. For two-sides hypotheses, significance level $\alpha = 0.05$ and $n = 36$, the critical value from Table 8 in Appendices is 0.221. We shall reject the null hypothesis H_0 when the test static D exceeds the critical value 0.221. The first step for calculation is to prepare the following table.

x	Freq	C. F.	F_S	z	F_T	$\|F_S - F_T\|$
70	2	2	0.0556	−2.14	0.0162	0.0394
75	2	4	0.1111	−1.43	0.0764	0.0347
79	2	6	0.1667	−0.86	0.1949	0.0282
83	3	9	0.2500	−0.29	0.3859	0.1359
86	2	11	0.3056	0.14	0.5557	0.2501
87	2	13	0.3611	0.29	0.6141	0.2530
88	5	18	0.5000	0.43	0.6664	0.1664
89	4	22	0.6111	0.57	0.7157	0.1046
91	6	28	0.7778	0.86	0.8051	0.0273
92	3	31	0.8611	1.00	0.8413	0.0198
95	1	32	0.8889	1.43	0.9236	0.0347
97	2	34	0.9444	1.71	0.9564	0.0120
98	2	36	1.0000	1.86	0.9686	0.0314
	36					

The first column has data, second column has frequency of data and third column has cumulative frequency. Now, $F_S(x)$ is obtained by dividing cumulative frequencies by the sample size, 36. For arithmetic mean 85 and standard deviation 7, we convert each observed value of x to a value of standard normal variable z through the relation $z = (x-85)/7$, given in fifth column. Form Table 1 in Appendices, we then obtain values of normal distribution integral for the limits from $-\infty$ to z. This value is $F_T(x)$, given in sixth column of the table.

The test statistic D can be computed algebraically or graphically. The possible values of $|F_S - F_T|$ are given in the last column in the table. This table shows that the value of test static D is 0.1547. Since the test static 0.253 is larger than the critical value 0.221, the null hypothesis is rejected. Hence the sample has not come from the specified distribution.

6. Problems and questions

1. What is sign test? For a randomly selected sample of 12 students, the scores obtained are as given in the following table. Apply the sign test with 6% significance level to decide if the median score of the population from which this sample is drawn is more than 15. Consider equal probability for plus and minus.

Student	Score	Student	Score
1	15	7	17
2	17	8	18
3	13	9	12
4	16	10	19
5	14	11	15
6	16	12	19

2. For a random sample of 15 patients, the results are as the following.

14.9	14.1	16.7	17.3	17.4	17.5	16.6	14.6
15.9	13.4	13.3	15.8	16.7	15.9	15.7	

Using the significance level 0.05, we wish to know if we can conclude on the basis of these data that the population mean is different from 15.

3. Using the significance level 0.06, for the following data, perform Wilcoxon test to conclude if the population mean is smaller than 17.

19	16	18	17	17	16	18	18	17	18
18	18	17	17	14	16	18	17	17	

4. A random sample of 12 male students from a rural junior high school and an independent random sample of 16 male students from an urban junior high school were given a test to measure their level of mental health. The result are given in the following table.

Urban	Rural	Urban	Rural
45	39	35	60
36	60	37	47
37	53	55	44
31	32	56	41
37	52	43	
48	57	36	
33	52	56	
35	42	51	

Use median test with 5% significance level to decide if urban and rural male junior high school students differ with respect to their level of mental health.

5. Let us design an experiment in order to assess the effects of prolonged inhalation of cadmium oxide. Fifteen laboratory animals served an experimental subjects, while ten similar animals served as controls. The variable of interest was hemoglobin level following the experiment. The results are shown in the following table. With the 5% significance level, we wish to know if we can conclude that median of X is smaller than median of Y.

Hemoglobin determinations (gm) for 25 laboratory animals

Exposed (X)	Unexposed (Y)	Exposed (X)	Unexposed (Y)
24.3	27.0	24.2	26.4
24.3	26.3	25.4	26.6
23.6	27.2	25.8	
26.4	27.6	26.6	
24.2	25.2	23.5	
26.5	26.3	25.4	
25.7	26.7	24.2	
25.8	25.2		

6. For 5% significance level, perform Kolmogorov-Smirnov test for the data given in the following table, in order to know if we may conclude that these data are not from a normally distributed population with arithmetic mean 70 and standard deviation 6.

65	74	70	67	58	77	82	67	82	76	68	66
70	71	62	67	82	70	71	67	67	82	58	77
74	65	68	70	70	67	62	71	66	68	71	76

7. Write short notes on the following:

(i) Sign-test

(ii) Wilcoxon test

(iii) Median test

(iv) Mann-Whitney U test

(v) t-test

(vi) Kolmogorov-Smirnov test

VIII. Analysis of Variance

The analysis of variance (abbreviated as ANOVA) is very useful technique for research in many fields. It plays very important role when there are cases having several samples. Though the significance of difference between the means of two samples can be judged with the help of z-test or t-test, but the difficulty arises when there are more than two samples at the same time. The ANOVA technique enables us to perform this simultaneous test and as such is considered to be an important tool for analysis. With the help of this technique, we can draw conclusions whether the samples have been drawn from the populations having the same mean.

The ANOVA technique is important in the context of all those situations where we want to compare more than two populations. For example, we want to compare the yields of crops from the use of several varieties of fertilizers, the smoking habits of several groups of workers in a factory, and so on. In such cases, we would not like to consider all possible combinations of two populations, as we have to make a large number of tests before reaching a decision. This whole process would consume a lot of time and money, and even then certain relationship may be left unidentified (particularly the interaction effects). Therefore, we would like to use the ANOVA technique where we can investigate the differences among the means of all the populations simultaneously.

1. ANOVA technique

Professor R.A. Fisher was the first person who coined the term 'variance' and developed a very elaborate theory about the ANOVA, explaining its usefulness for analyzing the samples. Later on, other scientists contributed in this field. The ANOVA is essentially a procedure for testing the difference among various groups of data. The essence of ANOVA is that the total amount of variance in a set of data in broken down into two parts: (i) the amount which can be assigned to the chance and (ii) the amount which can be assigned to some specified reasons. There may be variation between various samples and also within the items of a sample. ANOVA splits the variance for analytical purposes. Therefore, it is a method of analyzing the variance to which a response is subject into its various components, corresponding to various sources of variation.

1.1 Basic principle of ANOVA

The basic principle of ANOVA technique is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples. The F-ratio is expressed as

$$F = \frac{\text{Estimate of population variance between samples}}{\text{Estimate of population variance within samples}}$$

This value of F is to be compared to the F-limit for the given degrees of freedom[1]. When the calculated value of F is equal to or more than the F-limit given in table, we may say that there are significant differences between the sample means. On the other side, when calculated value of F is less than the table-value, the differences are considered to be insignificant or due to by chance.

2. One-way ANOVA technique

This is also known as the single factor ANOVA technique. It may be considered as a function of one independent variable. Under the one-way ANOVA, we consider only one factor and then observe that the reason for the said factor to be important is that several possible types of samples can occur within that factor. We then determine if there are differences within that factor. The technique involves the following steps:

(i) Suppose, we have k samples, so that $j = 1, 2, \ldots, k$ and $n_1, n_2, \ldots, n_k$ are, respectively, the number of items in the samples. Suppose, the values in the jth sample are X_{ij}, where $i = 1, 2, \ldots, n_j$.

(ii) Obtain the mean of each sample as

$$\overline{X}_j = \frac{\sum_i X_{ij}}{n_j} = \frac{X_{1j} + X_{2j} + \ldots + X_{n_j j}}{n_j} \qquad j = 1, 2, \ldots, k$$

[1]The F-value depends on the degrees of freedom of the numerator and the degrees of freedom of the denominator.

(iii) Calculate the mean of the sample-means as

$$\overline{\overline{X}} = \frac{\sum_j \overline{X}_j}{k} = \frac{\overline{X}_1 + \overline{X}_2 + \ldots + \overline{X}_k}{k}$$

(iv) Calculate the deviations of the sample means from the mean of the sample-means, square each of these deviations, multiply them by the number of items in the corresponding sample and then obtain their sum. This is known as the sum of the squares (abbreviated as SS) for variance between the samples (or SS between), expressed as

$$\text{SS between} = n_1(\overline{X}_1 - \overline{\overline{X}})^2 + n_2(\overline{X}_2 - \overline{\overline{X}})^2 + \ldots + n_k(\overline{X}_k - \overline{\overline{X}})^2$$

(v) Divide the 'SS between' by the degrees of freedom between the samples to obtain variance or mean square between the samples (or MS between), expressed as

$$\text{MS between} = \frac{\text{SS between}}{(k-1)}$$

where $(k-1)$ represents the degrees of freedom (d.f.) between samples.

(vi) Obtain the deviations of the values of the sample items for all the samples from corresponding means of the samples, calculate the squares of such deviations, and then obtain their sum. This sum is known as the sum of squares for variance within samples (or SS within), expressed as

$$\text{SS within} = \sum_{i=1}^{n_1}(X_{i1} - \overline{X_1})^2 + \sum_{i=1}^{n_2}(X_{i2} - \overline{X_2})^2 + \ldots + \sum_{i=1}^{n_j}(X_{ik} - \overline{X_k})^2$$

(vii) Divide the 'SS within' by the degrees of freedom within samples to obtain the variance or mean square within the samples (or MS within), expressed as

$$\text{MS within} = \frac{\text{SS within}}{(n-k)}$$

where $(n-k)$ represents degrees of freedom within samples. Here, $n = n_1 + n_2 + \ldots + n_k$, is the total number of items in all the samples.

(viii) Finally, the F-ratio is obtained as

$$F\text{-ratio} = \frac{\text{MS between}}{\text{MS within}}$$

This ratio is used to decide if the difference among various sample means is significant or not. For this purpose, we look into the table giving the F-value for the given degrees of freedom at different levels of significance. When the calculated value of F is less than the table-value of F, the difference is considered to be insignificant *i.e.,* due to chance. In case, the calculated value of F is either equal to or more than the table-value, the difference is considered to be significant and accordingly the conclusion may be drawn. The higher the calculated value of F is above the table-value, we are more definite and sure about the conclusions.

2.1 Short-cut for obtaining SS between and SS within

ANOVA can be performed with the help of the following short-cut method which is generally used in practice, as it is very convenient. The steps involved in the short-cut method are as the following.

(i) Calculate the sum T of all n values of individual items in all the samples as

$$T = \sum_{ij} X_{ij}$$

(ii) Calculate the correction factor T^2/n.

(iii) Calculate the square of all the items one by one and add the squares together. Subtract the correction factor from this sum and the result is known as the sum of squares for the total variance (SS total), expressed as

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n}$$

(iv) Calculate the square T_j^2 of each sample-total T_j and divide this square for each sample by the number of items in the concerned sample, and calculate the sum of the results so obtained. Subtract the correction factor from this total and the result is the sum of squares for variance between the samples (SS between), expressed as

$$\text{SS between} = \sum_{j} \frac{T_j^2}{n_j} - \frac{T^2}{n}$$

(v) The sum of the squares within the samples (SS within) can be obtained by subtracting the 'SS between' from the 'SS total' as

$$\text{SS within} = \text{SS total} - \text{SS between}$$

This is the short cut for obtaining SS between and SS within.

Exercise 1: Set up an analysis of variance for the following per acre production data for three varieties of wheat, each grown on 4 plots, and state if the variety differences are significant.

Per acre production data

	Variety of wheat		
Plot of land	P	Q	R
A	5	3	4
B	4	4	3
C	3	4	5
D	6	6	2

Solution: For the given set of data, we have the mean of each of these samples as the following.

$$\overline{X}_1 = \frac{5+4+3+6}{4} = 4.5$$

$$\overline{X}_2 = \frac{3+4+4+6}{4} = 4.25$$

$$\overline{X}_3 = \frac{4+3+5+2}{4} = 3.5$$

Mean of the sample means is

$$\overline{\overline{X}} = \frac{\overline{X}_1 + \overline{X}_2 + \overline{X}_3}{k} = \frac{4.5 + 4.25 + 3.5}{3} = 4.08$$

Now, we calculate SS between and SS within samples as the following.

$$\text{SS between} = n_1(\overline{X}_1 - \overline{\overline{X}})^2 + n_2(\overline{X}_2 - \overline{\overline{X}})^2 + n_3(\overline{X}_3 - \overline{\overline{X}})^2$$

$$= 4(4.5 - 4.08)^2 + 4(4.25 - 4.08)^2 + 4(3.5 - 4.08)^2 = 2.167$$

and

$$\text{SS within} = \sum_i (X_{1i} - \overline{X_1})^2 + \sum_i (X_{2i} - \overline{X_2})^2 + \sum_i (X_{ki} - \overline{X_k})^2$$

$$= [(5-4.5)^2 + (4-4.5)^2 + (3-4.5)^2 + (6-4.5)^2]$$
$$+[(3-4.25)^2 + (4-4.25)^2 + (4-4.25)^2 + (6-4.25)^2]$$
$$+[(4-3.5)^2 + (3-3.5)^2 + (5-3.5)^2 + (2-3.5)^2] = 14.75$$

Now, SS total can be obtained as

$$\text{SS total} = \text{SS between} + \text{SS within} = 2.17 + 14.75 = 16.92$$

Thus, we have

$$\text{MS between} = \frac{\text{SS between}}{(k-1)} = \frac{2.17}{3-1} = 1.09$$

and

$$\text{MS within} = \frac{\text{SS within}}{(n-k)} = \frac{14.75}{12-3} = 1.64$$

Finally,

$$F\text{-ratio} = \frac{\text{MS between}}{\text{MS within}} = \frac{1.09}{1.64} = 0.66$$

The F-value, $F(k-1, n-k)$, *i.e.*, $F(2,9)$ for 5% sifnificance level is 4.26. As the calculated value of F is 0.66 which is less than the table-value of 4.26 at 5% confidence level with d.f. being $v_1 = 2$ and $v_2 = 9$ and could have arisen due to variance ?. This analysis supports the null hypothesis of no difference means. We may therefore conclude that the difference in wheat output due to varieties and is just a matter of chance.

Short-cut for calculation of MS between and MS within

In this case, we first take the total of all the individual values of n items and call it as T, which is 49 in this case and $n = 12$. Thus, the correction factor $= T^2/n = 49 \times 49/12 = 200.08$. Now, SS total, SS between and SS within can be worked out as the following.

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n} = 5^2 + 4^2 + 3^2 + 6^2 + 3^2 + 4^2 + 4^2$$
$$+6^2 + 4^2 + 3^2 + 5^2 + 2^2 - \frac{49^2}{12} = 217 - 200.08 = 16.92$$

Now,

$$\text{SS between} = \sum_j \frac{T_j^2}{n_j} - \frac{T^2}{n} = \frac{18^2}{4} + \frac{17^2}{4} + \frac{14^2}{4} - \frac{49^2}{12}$$
$$= 202.25 - 200.08 = 2.17$$

and

$$\text{SS within} = \text{SS total} - \text{SS between} = 16.92 - 2.17 = 14.75$$

It may be noted that we get exactly the same result as we had obtained in the case of direct method.

2.2 Coding method

The coding method is based on the fact that the F-ratio remains unchanged when all the n items are either multiplied or divided by a common value or when a common value is added to or subtracted from each of the given n items. This method helps in reducing the big figures in magnitude by division or by subtraction. Consequently, the computation work is simplified without changing the F-ratio. This method should be used specially when the given figures are big. Once the given figures are converted with the help of coding, the other steps are the same, as discussed above. For example, the F-ratio remains unchanged for the data given in Table 1, in Table 2 and in Table 3.

Table 1. Parameters

	Action			
Group	$j = 1$	2	...	k
$i = 1$	X_{11}	X_{12}	...	X_{1k}
2	X_{21}	X_{22}	...	X_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	X_{m1}	X_{m2}	...	X_{mk}

Table 2. Parameters

	Action			
Group	$j = 1$	2	...	k
$i = 1$	$X_{11} - d$	$X_{12} - d$	...	$X_{1k} - d$
2	$X_{21} - d$	$X_{22} - d$	...	$X_{2k} - d$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	$X_{m1} - d$	$X_{m2} - d$	...	$X_{mk} - d$

Table 3. Parameters

	Action			
Group	$j = 1$	2	...	k
$i = 1$	X_{11}/d	X_{12}/d	...	X_{1k}/d
2	X_{21}/d	X_{22}/d	...	X_{2k}/d
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	X_{m1}/d	X_{m2}/d	...	X_{mk}/d

Exercise 2: Set up an analysis of variance for the following per acre production data for three varieties of wheat, each grown on 4 plots. Show that the SS between and SS within remain unchanged when we deduct some value, say 40, from each of these data.

Per acre production data

Plot of land	Variety of wheat		
	P	Q	R
A	48	46	44
B	47	43	45
C	45	42	41
D	46	44	47

Solution: The total all the individual values of $n = 12$ items is $T = 538$. Thus, the correction factor $= T^2/n = 538 \times 538/12 = 24120.33$. Now, SS total, SS between and SS within can be worked out as the following.

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n} = 48^2 + 47^2 + 45^2 + 46^2 + 46^2 + 43^2 + 42^2$$

$$+44^2 + 44^2 + 45^2 + 41^2 + 47^2 - \frac{538^2}{12} = 24170 - 24120.33 = 49.67$$

Now, we have

$$\text{SS between} = \sum_j \frac{T_j^2}{n_j} - \frac{T^2}{n} = \frac{186^2}{4} + \frac{175^2}{4} + \frac{177^2}{4} - \frac{538^2}{12}$$

$$= 241375 - 24120.33 = 17.17$$

and

$$\text{SS within} = \text{SS total} - \text{SS between} = 49.67 - 17.17 = 32.50$$

After subtraction of 40 from each of the values, we have

Per acre production data

Plot of land	Variety of wheat		
	P	Q	R
A	8	6	4
B	7	3	5
C	5	2	1
D	6	4	7

The total all the individual values of $n = 12$ items is $T = 58$. Thus, the correction factor $= T^2/n = 58 \times 58/12 = 280.33$. Now, SS total, SS between and SS within can be worked out as the following.

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n} = 8^2 + 7^2 + 5^2 + 6^2 + 6^2 + 3^2 + 2^2$$

$$+4^2 + 4^2 + 5^2 + 1^2 + 7^2 - \frac{58^2}{12} = 330 - 280.33 = 49.67$$

Now, we have

$$\text{SS between} = \sum_j \frac{T_j^2}{n_j} - \frac{T^2}{n} = \frac{26^2}{4} + \frac{15^2}{4} + \frac{17^2}{4} - \frac{58^2}{12}$$

$$= 297.5 - 280.33 = 17.17$$

and

$$\text{SS within} = \text{SS total} - \text{SS between} = 49.67 - 17.17 = 32.50$$

It obviously it shows that the values of SS between and SS within remain unchanged after subtraction of 40 from each term. Further, the ratio of SS between and SS within remains unchanged. Moreover, the F-ratio remains unchanged.

Exercise 3: Set up an analysis of variance for the following per acre production data for three varieties of wheat, each grown on 4 plots. Show that the F-ratio remains unchanged when we divide by some value, say 5, from each of these data.

Per acre production data

	Variety of wheat		
Plot of land	P	Q	R
A	45	40	35
B	55	45	50
C	35	40	55
D	30	25	35

Solution: The total all the individual values of $n = 12$ items is $T = 490$. Thus, the correction factor $= T^2/n = 490 \times 490/12 = 20008.33$. Now, SS total, SS between and SS within can be worked out as the following.

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n} = 45^2 + 55^2 + 35^2 + 30^2 + 40^2 + 45^2 + 40^2$$

$$+25^2 + 35^2 + 50^2 + 55^2 + 35^2 - \frac{490^2}{12} = 21008 - 20008.33 = 999.67$$

Now, we have

$$\text{SS between} = \sum_j \frac{T_j^2}{n_j} - \frac{T^2}{n} = \frac{165^2}{4} + \frac{150^2}{4} + \frac{175^2}{4} - \frac{490^2}{12}$$

$$= 200875 - 20008.33 = 79.17$$

and

$$\text{SS within} = \text{SS total} - \text{SS between} = 999.67 - 79.17 = 912.50$$

Thus, we have

$$\text{MS between} = \frac{\text{SS between}}{(k-1)} = \frac{79.17}{3-1} = 39.59$$

and

$$\text{MS within} = \frac{\text{SS within}}{(n-k)} = \frac{912.50}{12-3} = 101.39$$

Finally,

$$F\text{-ratio} = \frac{\text{MS between}}{\text{MS within}} = \frac{39.59}{101.39} = 0.39$$

After division by 5 each of the values, we have

Per acre production data

Plot of land	Variety of wheat		
	P	Q	R
A	9	8	7
B	11	9	10
C	7	8	11
D	6	5	7

The total all the individual values of $n = 12$ items is $T = 98$. Thus, the correction factor $= T^2/n = 98 \times 98/12 = 800.33$. Now, SS total, SS between and SS within can be worked out as the following.

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n} = 9^2 + 11^2 + 7^2 + 6^2 + 8^2 + 9^2 + 8^2$$

$$+5^2 + 7^2 + 10^2 + 11^2 + 7^2 - \frac{98^2}{12} = 840 - 800.33 = 39.67$$

Now, we have

$$\text{SS between} = \sum_j \frac{T_j^2}{n_j} - \frac{T^2}{n} = \frac{33^2}{4} + \frac{30^2}{4} + \frac{35^2}{4} - \frac{98^2}{12}$$

$$= 803.5 - 800.33 = 3.17$$

and

$$\text{SS within} = \text{SS total} - \text{SS between} = 39.67 - 3.17 = 36.50$$

Thus, we have

$$\text{MS between} = \frac{\text{SS between}}{(k-1)} = \frac{3.17}{3-1} = 1.59$$

and

$$\text{MS within} = \frac{\text{SS within}}{(n-k)} = \frac{36.50}{12-3} = 4.06$$

Finally,

$$F\text{-ratio} = \frac{\text{MS between}}{\text{MS within}} = \frac{1.59}{4.06} = 0.39$$

It obviously it shows that the F-ratio remains unchanged after division by 5 each of the terms. Further, the ratio of SS between and SS within also remains unchanged.

3. Two-way ANOVA technique

The two-way ANOVA technique is used when the data are classified on the basis of two factors. It may be considered as a function of two independent variables. For example, the agriculture output may be classified on the basis of different varieties of seeds and also on the basis of different varieties of fertilizers used. A business firm may have its sale data classified on the basis of different salesmen and also on the basis of different regions considered. Such a two-way design may have (i) repeated values of each factor or (ii) may not have repeated values. The ANOVA technique is little different in case of repeated measurements where we also compute the interaction variation. We shall here discuss about both the cases.

3.1 When repeated values are not there

The various steps involved in this case are as the following.

(i) Calculate the sum T of all n values of individual items in all samples as

$$T = \sum_{ij} X_{ij}$$

(ii) Calculate the correction factor T^2/n.

(iii) Calculate the square of all the items one by one and add the squares together. Subtract the correction factor from this total and the result is known as the sum of squares for the total variance (SS total), expressed as

$$\text{SS total} = \sum_{ij} X_{ij}^2 - T^2/n$$

(iv) Calculate the sum of different columns, T_j, then obtain the square of T_j, divide such squared value by the number n_j of items in the concerned column and take the sum of the result so obtained. Finally, subtract the correction-factor from this sum to obtain the sum of squares of deviations for variance between columns (SS between columns) as

$$\text{SS between columns} = \sum_j \frac{T_j^2}{n_j} - \frac{T^2}{n}$$

(v) Calculate the sum of different rows, T_i, then obtain the square of T_i, divide such squared value by the number n_i of items in the concerned row and take the sum of the result so obtained. Finally, subtract the correction-factor from this sum to obtain the sum of squares of deviations for variance between rows (SS between rows) as

$$\text{SS between rows} = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n}$$

(vi) Sum of squares of deviations for residual or error variance can be obtained by subtracting the 'SS between columns' and 'SS between rows' from the 'SS total'.

$$\text{SS for residual} = \text{SS total} - [\text{SS between columns} + \text{SS between rows}]$$

(vii) When c is the number of columns and r the number of rows, the degrees of freedom (d.f.) can be obtained as the following.

d.f. for total variance $= (c.r - 1)$

d.f. for variance between columns $= (c - 1)$

d.f. for variance between rows $= (r - 1)$

d.f. for residual variance $= (c - 1)(r - 1)$

(viii) Now, we have

$$\text{MS between columns} = \frac{\text{SS between columns}}{c-1}$$

$$\text{MS between rows} = \frac{\text{SS between rows}}{r-1}$$

$$\text{MS for residual} = \frac{\text{SS for residual}}{(r-1)(c-1)}$$

(ix) Finally,

$$F\text{-ratio for columns} = \frac{\text{MS between columns}}{\text{MS for residual}}$$

$$F\text{-ratio for rows} = \frac{\text{MS between rows}}{\text{MS for residual}}$$

(x) These calculated F-ratios are compared with the $F([c-1],[r-1][c-1])$ and $F([r-1],[r-1][c-1])$, respectively.

These F-ratios are used to decide if the impacts of the parameters are significant or not. When the calculated value of F is less than the table-value of F, the difference is considered to be insignificant *i.e.*, due to chance. In case, the calculated value of F is either equal to or more than the table-value, the difference is considered to be significant and accordingly the conclusion may be drawn. The higher the calculated value of F is above the table-value, we are more definite and sure about the conclusions.

Exercise 4: Set up an ANOVA analysis for the following two-way design results.

Per acre production data

Plot of land	Variety of wheat		
	P	Q	R
A	8	9	7
B	6	5	4
C	8	6	4
D	9	5	3

Also state whether variety differences are significant at 5% level.

Solution: For this two-way design we shall proceed in the following manner.

(i) We first calculate the total of all $n = 12$ items

$$T = 8 + 6 + 8 + 9 + 9 + 5 + 6 + 5 + 7 + 4 + 4 + 3 = 74$$

(ii) The correction factor is

$$\text{Correction-factor} = \frac{T^2}{n} = \frac{74^2}{12} = 456.33$$

(iii) We have

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n} = 8^2 + 6^2 + 8^2 + 9^2 + 9^2 + 5^2 + 6^2 + 5^2$$
$$+7^2 + 4^2 + 4^2 + 2^2 - \frac{74^2}{12} = 502 - 456.33 = 45.67$$

Sums of the columns are 31, 25, 18 and therefore,

$$\text{SS between columns} = \sum_j \frac{T_j^2}{n_j} - \frac{T^2}{n} = \frac{31^2}{4} + \frac{25^2}{4} + \frac{18^2}{4} - \frac{74^2}{12}$$
$$= 477.5 - 456.33 = 21.67$$

Sums of the rows are 24, 15, 18, 17 and therefore,

$$\text{SS between rows} = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n} = \left[\frac{24^2}{3} + \frac{15^2}{3} + \frac{18^2}{3} + \frac{17^2}{3}\right] - \frac{74^2}{12}$$
$$= 471.33 - 456.33 = 15$$

(iv) Therefore, we have

$$\text{SS residual} = \text{SS total} - (\text{SS between columns} + \text{SS between rows})$$
$$= 45.67 - (21.17 + 15) = 9.5$$

(v) Now, we have

$$\text{MS between columns} = \frac{\text{SS between columns}}{c - 1} = \frac{21.17}{3 - 1} = 10.59$$

$$\text{MS between rows} = \frac{\text{SS between rows}}{r - 1} = \frac{15}{4 - 1} = 5$$

$$\text{MS for residual} = \frac{\text{SS for residual}}{(r - 1)(c - 1)} = \frac{9.5}{2 \times 3} = 1.58$$

(vi) Finally,

$$F\text{-ratio for columns} = \frac{\text{MS between columns}}{\text{MS for residual}} = \frac{10.59}{1.58} = 6.70$$

$$F\text{-ratio for rows} = \frac{\text{MS between rows}}{\text{MS for residual}} = \frac{5}{1.58} = 3.16$$

(vii) Now, at 5% level, we have

$$F([r-1],[r-1][c-1]) = F(2,6) = 5.14$$

and

$$F([c-1],[r-1][c-1]) = F(3,6) = 4.76$$

Since the calculated value of F for the columns is larger than the corresponding table-value, the differences due to varieties are significant. Since the calculated value of F for the rows is smaller than the corresponding table-value, the differences due to plots are insignificant.

3.2 When repeated values are there

In case of a two-way design with repeated values for all the categories, we consider the following table where the repeated values X and Y are given. The repetition may be more than two times.

Repeated values X and Y

	Action			
Group	$j=1$	2	$\dots$	k
$i=1$	X_{11} Y_{11}	X_{12} Y_{12}	$\dots$ $\dots$	X_{1k} Y_{1k}
2	X_{21} Y_{21}	X_{22} Y_{22}	$\dots$ $\dots$	X_{2k} Y_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	X_{m1} Y_{m1}	X_{m2} Y_{m2}	$\dots$ $\dots$	X_{mk} Y_{mk}

The following steps are taken. Suppose, there are c columns and r rows so that $c = k$ and $r = m$.

(i) We first calculate the sum of all $n = 2cr$ items

$$T = \sum_{i,j} X_{ij} + \sum_{i,j} Y_{ij}$$

Here, the factor 2 is introduced because of two times repetition.

(ii) The correction factor is T^2/n.

(iii) We have

$$\text{SS total} = \sum_{ij} (X_{ij}^2 + Y_{ij}^2) - \frac{T^2}{n}$$

$$\text{SS between columns} = \frac{1}{2m} \sum_{j=1}^{k} \Big[\sum_{i=1}^{m} (X_{ij} + Y_{ij}) \Big]^2 - \frac{T^2}{n}$$

$$\text{SS between rows} = \frac{1}{k} \sum_{i=1}^{m} \Big[\Big\{ \sum_{j=1}^{k} X_{ij} \Big\}^2 + \Big\{ \sum_{j=1}^{k} Y_{ij} \Big\}^2 \Big] - \frac{T^2}{n}$$

$$\text{SS within samples} = \sum_{ij} \Big[(X_{ij} - V_{ij})^2 + (Y_{ij} - V_{ij})^2 \Big]$$

where $V_{ij} = (X_{ij} + Y_{ij})/2$ is the average over all the repetitions. Here, repetition is two times.

$$\text{SS for interaction} = \text{SS total} - [\text{SS between columns}$$

$$+\text{SS between rows} + \text{SS within samples}]$$

(iv) The degrees of freedom (d.f.) can be obtained as the following.

d.f. for total variance $= (2cr - 1) = n - 1$

d.f. for variance between columns $= (c - 1)$

d.f. for variance between rows $= (r - 1)$

d.f. for variance within samples $= 2cr - cr = cr$

d.f. for interaction $= (2cr - 1) - [(c - 1) + (r - 1) + cr]$
$= (c - 1)(r - 1)$

(v) Now, we have

$$\text{MS between columns} = \frac{\text{SS between columns}}{c-1}$$

$$\text{MS between rows} = \frac{\text{SS between rows}}{r-1}$$

$$\text{MS for interaction} = \frac{\text{SS for interaction}}{(r-1)(c-1)}$$

$$\text{MS within samples} = \frac{\text{SS within samples}}{cr}$$

(vi) Finally,

$$F\text{-ratio for columns} = \frac{\text{MS between columns}}{\text{MS within samples}}$$

$$F\text{-ratio for rows} = \frac{\text{MS between rows}}{\text{MS within samples}}$$

$$F\text{-ratio for interaction} = \frac{\text{MS for interaction}}{\text{MS within samples}}$$

(vii) These calculated F-ratios are compared with the $F([c-1], cr)$, $F([r-1], cr)$ and $F([r-1][c-1], cr)$, respectively.

These F-ratios are used to decide if the impacts of the parameters are significant or not. When the calculated value of F is less than the table-value of F, the difference is considered to be insignificant *i.e.,* due to chance. In case, the calculated value of F is either equal to or more than the table-value, the difference is considered to be significant and accordingly the conclusions may be drawn. The higher the calculated value of F is above the table-value, we are more definite and sure about the conclusions. All these steps can be understood with the help of the following exercise.

Exercise 5: Apply the two-way ANOVA technique for the following information about the three drugs (P, Q, R) testing to decide about the effectiveness for reduction of blood pressure for four different groups (A, B, C, D) of people. Do the drugs affect differently? Are the different groups of people affected differently? Is the interaction term significant? Answer these questions taking a significant level of 5%.

Blood pressure reduction in mm of Hg

Group of people	Drug		
	P	Q	R
A	12	9	13
	14	8	12
B	14	8	13
	15	9	12
C	13	12	10
	14	11	9
D	15	14	13
	16	12	11

Solution: We shall calculate two-way NOVA technique as the following.

(i) We first calculate the total of all $n = 24$ items

$$T = \sum_{ij} \left[X_{ij} + Y_{ij}\right] = 12 + 14 + 9 + 8 + 13 + 12 + 14 + 15 + 8 + 9 + 13$$

$$+12 + 13 + 14 + 12 + 11 + 10 + 9 + 15 + 16 + 14 + 12 + 13 + 11 = 289$$

(ii) The correction factor is

$$\text{Correction-factor} = \frac{T^2}{n} = \frac{289^2}{24} = 3480.04$$

(iii) We have

$$\text{SS total} = \sum_{ij} \left[X_{ij}^2 + Y_{ij}^2\right] - \frac{T^2}{n} = [12^2 + 14^2 + 9^2 + 8^2 + 13^2 + 12^2 +$$

$$14^2 + 15^2 + 8^2 + 9^2 + 13^2 + 12^2 + 13^2 + 14^2 + 12^2 + 11^2 + 10^2 + 9^2 +$$

$$15^2 + 16^2 + 14^2 + 12^2 + 13^2 + 11^2] - \frac{289^2}{24} = 3599 - 3480.04 = 118.96$$

Sums of the columns are 113, 83 and 93, and

$$\text{SS between columns} = \frac{1}{2r} \sum_{j=1}^{4} \Big[\sum_{i}^{4} X_{ij} + \sum_{i}^{4} Y_{ij}\Big]^2 - \frac{T^2}{n}$$

$$= \frac{113^2}{8} + \frac{83^2}{8} + \frac{93^2}{8} - \frac{289^2}{24} = 3538.38 - 3480.04 = 58.34$$

Sums of the rows are 34, 34, 35, 36, 35, 34, 42 and 39, and

$$\text{SS between rows} = \frac{1}{k}\sum_{i=1}^{m}\Big[\Big\{\sum_{j=1}^{k} X_{ij}\Big\}^2 + \Big\{\sum_{j=1}^{k} Y_{ij}\Big\}^2\Big] - \frac{T^2}{n}$$

$$= \Big[\frac{34^2}{3} + \frac{34^2}{3} + \frac{35^2}{3} + \frac{36^2}{3} + \frac{35^2}{3} + \frac{34^2}{3} + \frac{42^2}{3} + \frac{39^2}{3}\Big] - \frac{289^2}{24}$$

$$= 3499.67 - 3480.04 = 19.63$$

$$\text{SS within samples} = (12-13)^2 + (14-13)^2 + (14-14.5)^2 +$$

$$(15-14.5)^2 + (13-13.5)^2 + (14-13.5)^2 + (15-15.5)^2 + (16-15.5)^2 +$$

$$(9-8.5)^2 + (8-8.5)^2 + (8-8.5)^2 + (9-8.5)^2 + (12-11.5)^2 +$$

$$(11-11.5)^2 + (14-13)^2 + (12-13)^2 + (13-12.5)^2 + (12-12.5)^2 +$$

$$(13-12.5)^2 + (12-12.5)^2 + (10-9.5)^2 + (9-9.5)^2 + (13-12)^2 +$$

$$(11-12)^2 = 10.50$$

$$\text{SS for interaction} = 118.96 - [58.34 + 19.63 + 10.50] = 30.49$$

(iv) The degrees of freedom (d.f.) can be obtained as the following.

d.f. for total variance $= n - 1 = 24 - 1 = 23$
d.f. for variance between columns $= (c-1) = 3 - 1 = 2$
d.f. for variance between rows $= (r-1) = 4 - 1 = 3$
d.f. for variance within sample $= n - cr = 24 - 12 = 12$
d.f. for interaction $= (c-1)(r-1) = 6$

(iv) We have

$$\text{MS between columns} = \frac{\text{SS between columns}}{\text{d.f.}} = \frac{58.34}{3-1} = 29.17$$

$$\text{MS between rows} = \frac{\text{SS between rows}}{\text{d.f.}} = \frac{19.63}{4-1} = 6.54$$

$$\text{MS for interaction} = \frac{\text{SS for interaction}}{\text{d.f.}} = \frac{30.49}{6} = 5.08$$

$$\text{MS within samples} = \frac{\text{SS within samples}}{\text{d.f.}} = \frac{10.50}{24-12} = 0.88$$

(v) Finally, we have

$$F\text{-ratio between columns} = \frac{29.17}{0.88} = 33.15$$

$$F\text{-ratio between rows} = \frac{6.54}{0.88} = 7.43$$

$$F\text{-ratio for interaction} = \frac{5.08}{0.88} = 5.77$$

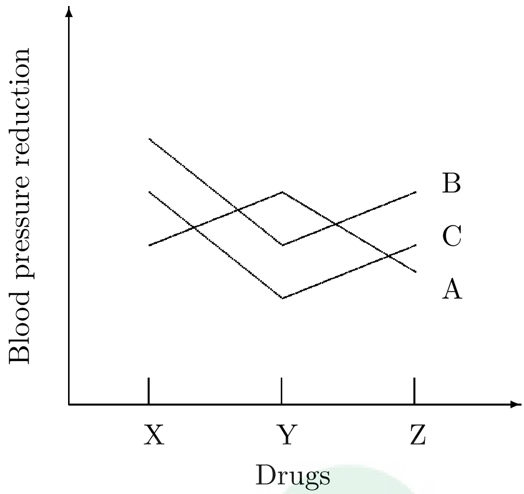
For 5% confidence level, we have

$F(2, 12) = 3.89$ $\quad$ $F(3, 12) = 3.49$ $\quad$ $F(6, 12) = 3.00$

Since all the three calculated F-ratios are larger than corresponding table-value, it shows that all the three F-ratios are significant at 5% significant level. It means that the drugs act differently, different groups of people are affected differently and the interaction term is significant.

3.3 Graphic method of studying interaction in a two-way design

Interaction can be studied in two-way design with repeated measurements through graphic method also. For such a graph, we plot one of the factors on the horizontal axis. Then we plot the averages for all the samples on the graph and connect the averages for each variety of the other factor by a distinct mark (or a coloured line). When the connecting lines do not cross over each other, then the graph indicates that there is no interaction. On the other side, when the lines cross over each other, they indicate definite interaction or inter-relation between the two factors. Figure shows such graph for some data to see if there is any interaction between the two factors *viz.*, the drugs and the groups of people.



The graph indicates that there is interaction as the connecting lines for groups of people cross over each other. We find that A and B are affected very similarly, but C is affected differently. The highest reduction in blood pressure in case of C is with drug Y and the lowest reduction is with drug Z, whereas the highest reduction in blood pressure in cases of A and B is with drug X and the lowest reduction with drug Y. Hence, there is definite inter-relation between the drugs and the groups of people, and cannot make any strong statements about drugs unless one also qualifies his/her conclusions by stating which group of people one is dealing with. In such a situation, performing F-test is meaningless. But, when the lines do not cross over each other (and remain more or less identical), then there is no interaction or the interaction is not considered a significantly large value, in which case one should proceed to test the main effects, drugs and people in the given case, as stated earlier.

4. ANOVA in a Latin square design

The ANOVA technique in case of Latin square design remains more or less the same as in case of a two-way design, excepting the fact that the variance is splitted into four pairs. Moreover, the number of columns c, number of rows r and number of varieties v are equal, *i.e.*, $c = r = v$. Various steps used are as the following.

(i) Calculate the sum T of all n values of individual items in all samples

$$T = \sum_{ij} X_{ij}$$

(ii) Calculate the correction factor T^2/n.

(iii) We have

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n}$$

(iv) Calculate the sum of different columns, T_j, then obtain the square of T_j, divide such squared value by the number n_j of items in the concerned column and take the sum of the result so obtained. Finally, subtract the correction-factor from this sum to obtain the sum of squares of deviations for variance between columns (SS between columns) as

$$\text{SS between columns} = \sum_{j} \frac{T_j^2}{n_j} - \frac{T^2}{n}$$

(v) Calculate the sum of different rows, T_i, then obtain the square of T_i, divide such squared value by the number n_i of items in the concerned row and take the sum of the result so obtained. Finally, subtract the correction-factor from this sum to obtain the sum of squares of deviations for variance between rows (SS between rows) as

$$\text{SS between rows} = \sum_{i} \frac{T_i^2}{n_i} - \frac{T^2}{n}$$

(vi) Further, we have

$$\text{SS between varieties} = \sum_{v} \frac{T_v^2}{n_v} - \frac{T^2}{n}$$

and

$$\text{SS residual} = \text{SS total} - [\text{SS between columns} + \text{SS between rows} + \text{SS between varieties}]$$

(vii) The degrees of freedom (d.f.) can be obtained as the following.

d.f. for variance between columns $= (c - 1)$

d.f. for variance between rows $= (r - 1)$

d.f. for variance between varieties $= (v - 1)$

d.f. for total variance $= (2cr - 1) = n - 1$

d.f. for variance residual $= (c - 1)(c - 2)$

(viii) Now, we have

$$\text{MS between columns} = \frac{\text{SS between columns}}{(c-1)}$$

$$\text{MS between rows} = \frac{\text{SS between rows}}{(r-1)}$$

$$\text{MS between varieties} = \frac{\text{SS between varieties}}{(v-1)}$$

$$\text{MS residual} = \frac{\text{SS residual}}{(c-1)(c-2)}$$

(ix) Further, we have

$$F\text{-ratio between columns} = \frac{\text{MS between columns}}{\text{MS for residual}}$$

$$F\text{-ratio between rows} = \frac{\text{MS between rows}}{\text{MS for residual}}$$

$$F\text{-ratio between varieties} = \frac{\text{MS between varieties}}{\text{MS for residual}}$$

(x) These calculated F-ratios are compared with the $F([c-1],[(c-1)(c-2)])$, $F([r-1],[(c-1)(c-2)])$ and $F([v-1],[(c-1)(c-2)])$, respectively.

These F-ratios are used to decide if the impacts of the parameters are significant or not. When the calculated value of F is less than the table-value of F, the difference is considered to be insignificant *i.e.,* due to by chance. In case, the calculated value of F is either equal to or more than the table-value, the difference is considered to be significant and accordingly the conclusions may be drawn. The higher the calculated value of F is above the table-value, we are more definite and sure about the conclusions. All these steps can be understood with the help of the following exercise.

Exercise 6: Analyze and interpret the following statistics about the output of wheat per field obtained as a result of experiment conducted to test four varieties of wheat, *viz.*, A, B, C, D under a Latin square design.

C 23	B 21	A 19	D 18
A 25	D 24	C 20	B 19
B 21	A 16	D 18	C 22
D 15	C 19	B 22	A 16

Solution: Using the coding method, we subtract 20 from the numbers given in each of the small squares and obtain the coded figures as the following.

Rows	Columns				Row total
	1	2	3	4	
1	C 3	B 1	A −1	D −2	1
2	A 5	D 4	C 0	B −1	8
3	B 1	A −4	D −2	C 2	−3
4	D −5	C −1	B 2	A −4	−8
Column total	4	0	−1	−5	T = −2

Squaring these coded figures in various columns and rows, we get

Rows	Columns				Row total
	1	2	3	4	
1	C 9	B 1	A 1	D 4	15
2	A 25	D 16	C 0	B 1	42
3	B 1	A 16	D 4	C 4	25
4	D 25	C 1	B 4	A 16	46
Column total	60	34	9	25	T = 128

(i) The total of all values of individual items in all the samples is

$$T = \sum_{ij} X_{ij} = -2$$

(ii) The correction factor is

$$\text{Correction factor} = \frac{T^2}{n} = \frac{(-2)^2}{16} = 0.25$$

(iii) Now, we have

$$\text{SS total} = \sum_{ij} X_{ij}^2 - \frac{T^2}{n} = 9 + 25 + 1 + 25 + 1 + 16 + 16 + 1 + 1$$

$$+0 + 4 + 4 + 4 + 1 + 4 + 16 - \frac{(-2)^2}{16} = 128 - 0.25 = 127.75$$

Sums of the columns are 4, 0, −1, −5, and therefore,

$$\text{SS between columns} = \sum_{j} \frac{T_j^2}{n_j} - \frac{T^2}{n} = \left[\frac{4^2}{4} + \frac{0^2}{4} + \frac{(-1)^2}{4} + \frac{(-5)^2}{4}\right]$$

$$-0.25 = 10.25$$

Sums of the rows are 1, 8, −3, −8, and therefore,

$$\text{SS between rows} = \sum_{i} \frac{T_i^2}{n_i} - \frac{T^2}{n} = \left[\frac{1^2}{4} + \frac{8^2}{4} + \frac{(-3)^2}{4} + \frac{(-8)^2}{4}\right] - 0.25$$

$$= 34.25$$

For calculating SS for variance between varieties, we first rearrange the coded data in the following form.

Variety of wheat	Yield in different part of field				Total
	I	II	III	IV	T_v
A	5	−4	−1	−4	−4
B	1	1	2	−1	3
C	3	−1	0	2	4
D	−5	4	−2	−2	−5

Now, the SS for variance between varieties is

$$\text{SS between varieties} = \sum_i \frac{T_v^2}{n_v} - \frac{T^2}{n} = \left[\frac{(-4)^2}{4} + \frac{3^2}{4} + \frac{4^2}{4} + \frac{(-3)^2}{4}\right]$$

$$-\frac{(-2)^2}{16} = 16.5 - 0.25 = 16.25$$

Therefore, the sum of the squares for residual variance is

$$\text{SS for residual variance} = 127.75 - (10.25 + 34.25 + 16.25) = 67$$

(iv) The degrees of freedom (d.f.) can be obtained as the following.

d.f. for variance between columns = $(c-1)$ = 4 - 1 = 3

d.f. for variance between rows = $(r-1)$ = 4 - 1 = 3

d.f. for variance between varieties = $(v-1)$ = 4 - 1 = 3

d.f. for total variance = $(n-1)$ = 16 - 1 = 15

d.f. for residual variance = $(c-1)(c-2)$ = (4-1) (4-2) = 6

(v) We have

$$\text{MS between columns} = \frac{\text{SS between columns}}{\text{d.f.}} = \frac{10.25}{4-1} = 3.42$$

$$\text{MS between rows} = \frac{\text{SS between rows}}{\text{d.f.}} = \frac{34.25}{4-1} = 11.42$$

$$\text{MS between varieties} = \frac{\text{SS between varieties}}{\text{d.f.}} = \frac{16.25}{4-1} = 5.42$$

$$\text{MS for residual} = \frac{\text{SS for residual}}{\text{d.f.}} = \frac{67}{6} = 11.17$$

(vi) Further, we have

$$F\text{-ratio between columns} = \frac{\text{MS between columns}}{\text{MS for residual}} = \frac{3.42}{11.17} = 0.31$$

$$F\text{-ratio between rows} = \frac{\text{MS between rows}}{\text{MS for residual}} = \frac{11.42}{11.17} = 1.02$$

$$F\text{-ratio between varieties} = \frac{\text{MS between varieties}}{\text{MS for residual}} = \frac{5.42}{11.17} = 0.49$$

For 5% significance level, we have $F(3, 6) = 4.76$. It shows that the differences between columns, between rows and between varieties are insignificant.

5. Problems and questions

1. Discuss about the one-way ANOVA technique.

2. For one-way ANOVA analysis for the following data, show that the SS between and SS within remain unchanged when we deduct 10 from each of these data.

	Parameter 1		
Parameter 2	P	Q	R
A	13	15	14
B	12	14	16
C	14	12	10

3. Set up one-way ANOVA for the data for per acre production given in the following table, and state if the variety differences are significant for 5% significance level.

	Variety of wheat		
Plot of land	P	Q	R
A	3	5	2
B	4	4	3
C	5	6	5

4. Apply the two-way ANOVA technique for the following information about the three drugs (P, Q, R) testing to decide about the effectiveness for reduction of blood pressure for three different groups (A, B, C) of people. Do the drugs affect differently? Are the different groups of people affected differently? Is the interaction

term significant? Answer these questions taking a significant level of 5%.

Blood pressure reduction in mm of Hg

Group of people	Drug		
	P	Q	R
A	11	10	13
	13	9	11
B	13	9	12
	14	8	11
C	12	11	10
	13	10	9

5. Discuss about the two-way ANOVA technique.

6. Discuss about the ANOVA technique in context of two-step design when repeated values are not there.

7. Discuss about the ANOVA technique in context of two-step design when repeated values are there.

8. Write short notes on the following:

   (i) One-way ANOVA

   (ii) Two-way ANOVA

   (iii) Coding method

   (iv) ANOVA in a Latin square design

IX. Simulation

Pertaining to our real life, there are many problems which cannot be represented mathematically due to stochastic nature, complexity or conflicting ideas required to describe them. When all other methods fail, for such problems, simulation is often used. The simulation technique has long been used by the analysts and designers in physical sciences and has now become an important tool for dealing with the complicated problems. The first important application of simulation was probably was made by John Von Neumann and Stanislaw Ulam for determining the complicated behaviour of neutrons in a nuclear shielding problem being too complex for mathematical analysis.

After getting the remarkable success of this technique on neutron problem, it became more popular and is being used in several areas. In 1950's, the development of digital computer further increased the importance of the simulation technique. In the present chapter, we shall discuss the process of simulation and necessary tools to perform such analysis. A special emphasis is given to the Monte-Carlo method of simulation. Some simple examples are discussed to explain the Monte-Carlo technique. To obtain the reliable results, the use of computer is very essential in the technique. However, for easy demonstration of simulation technique, the numerical examples are solved by the pocket calculator only.

1. Introduction

In fact, simulation is a representative model for a real situation. While visiting some trade-fair or exhibition, we often find a number of simulated environments therein. In laboratories, we often perform a number of experiments on simulated models to predict behaviour of the real system under true environments. The combination of computing and simulation has also resulted in the production of TV games. Players interrupt the way of computer program moves, various images around the screen from a keyboard or hand-held controller. The computer incorporated their responses into these movements in accordance with the rules of the particular game. Incidentally, such programs make use of random numbers. For example, the random numbers may be used to find deflection of tennis balls, etc. Before proceeding further, let us define the term simulation in more suitable forms. Following are some of the definitions:

Definition 1. Simulation is a representation of reality through the use of model or other device which reacts in the same manner as the reality under a given set of conditions.

Definition 2. Simulation is an use of system model that has designed the characteristics of the reality in order to produce the essence of actual operation.

1.1 Types of simulations

Simulations are mainly classified into two categories:

(i) Analogue simulation (or environment simulation): The simple examples such as children's cyclic park with various signals, crossings in an exhibition, testing of an aircraft model in a wind tunnel, are examples of the analogue simulation.

(ii) Computer simulation (or system simulation): For the complex and intricate problems of managerial decision making, the analogue simulation may not be applicable, and the actual experimentation with the system may be uneconomical also. Under these situations, the complex system is formulated into a mathematical model for which a computer program is developed, and then the problem is solved with the help of high speed electronic computer. Such type of simulation is known as a computer simulation or system simulation.

1.2 Classification of simulation models

The simulation models can be classified into the following four categories:

(a) Deterministic models: In these models, input and output variables are not permitted to be random variables and models are described by exact functional relationship.

(b) Stochastic models: In these models, at least one of the variables or functional relationship is expressed by probability functions.

(c) Static models: In these models, we do not take variable time into consideration.

(d) Dynamic models: In these models, we account for the time varying interaction.

2. Need of simulation

In the operations research, various types of managerial decision making problems can be solved by (i) scientific method, (ii) analytical method, and (iii) iterative method. But each method has its drawbacks and limitations as discussed in the following.

2.1 Drawbacks of scientific method

The steps of scientific method have the following limitations and difficulties.

(i) It may be either impossible or extremely costly to observe certain processes in the real life situations.

(ii) The observed system may be so complex that it may be impossible to describe it in terms of a set of mathematical equations.

(iii) Even though the mathematical model can be formulated to describe system under study, a straight forward analytical solution may not be available. For example, such situations may arise in complex queuing problems, job-shop problems, high-order differential equations, complicated stochastic models, multi-integral problems, etc.

(iv) It may be either impossible or very costly to perform validating experiments on mathematical models describing the system.

Thus, on account of these drawbacks of scientific method cannot be used to solve complex managerial decision-making problems.

2.2 Drawbacks of analytical method

Analytical technique used in dynamic programming, queuing theory, network models, etc., are not sufficient to tackle all the important managerial problems, requiring data analysis, due to following limitations.

(i) Dynamic programming models can be used to determine optimal strategies taking into account the uncertainties and can analyze multi-period planning problems. But, still it has its own shortcoming. Dynamic programming models can be used to tackle very simple situations involving a very few variables. If the number of state variables becomes larger, the computation work becomes quite complex and difficult.

(ii) Similar limitations also hold good for other mathematical techniques, dynamic stochastic models such as inventory and waiting line situations. Only small scale systems are amenable to these models. But, by making a number of assumptions the systems are simplified to such an extent that in many cases, the results thus obtained are only rough approximations.

2.3 Drawbacks of iterative method

(i) In linear programming models, we assume that data do not change over the entire planning horizon. This is one time decision process and assumes average values for the decision variables. If the planning horizon is long, say 15 years, the multi-period linear programming model may deal with the yearly averaged data, but will not take into account the variations over the months and weeks. Consequently, month to month and week to week operations are left implicit.

(ii) Other important limitations of linear programming is that it assumes that data should be known with certainty. In many real situations, the uncertainties about the data are such that they cannot be ignored. In case the uncertainty relates to only a few variables, the sensitivity analysis can be used to determine its effect on the decision. But, in the situations, where uncertainty pervades the entire model, the sensitivity analysis may become too cumbersome and computationally difficult to determine the impact of uncertainty on recommended plan.

From the above stated drawbacks, we conclude that whenever the characteristics like uncertainty, complexity, dynamic interaction between the decision and subsequent event and the need to develop detailed procedures and finely divided time intervals, all combined together in one situation, then model becomes too complex to be solved by any of the techniques of mathematical programming and probabilistic models.

Then such complex model must be analyzed by some other kind of quantitative technique, which may give accurate and reliable results. Many new techniques are investigated so far, but among all, the best available is the simulation.

In general, the simulation technique is a dependable tool in situations where mathematical analysis is either too complex to too expensive.

3. Limitation of simulation technique

Although many operations research analysts consider the simulation as a method of last resort and use it only when all other techniques fail. When the problem can be well represented by a mathematical model, the analytical method is considered to be more economical, accurate and reliable. But, the case of very large and complex problems, simulation may suffer the similar drawbacks as other mathematical models. The limitations of the simulation technique may be briefly mentioned as the following.

(i) Optimum results cannot be produced by simulation. Since the model mostly deals with uncertainties, the results of simulation are only reliable approximations involving statistical errors.

(ii) The another difficulty lies in the quantification of the variables. In many situations, it is not possible to quantify all the variables which affect the behaviour of the system.

(iii) In very large and complex problems, it becomes difficult to make the computer program on account of large number of variables and the involved inter-relationships among them. The number of variables may be too large and may exceed beyond the capacity of available computers.

(iv) For the problem requiring use of computer, the simulation is by no means a less costly method of analysis. Thus, simulation is comparatively costlier and time consuming method in many situations.

(v) Other important limitations of them are from too much tendency to rely on the simulation models. Consequently, this technique is also applied to some simple problems which can otherwise be solved by more appropriate techniques of mathematical programming.

4. Phases of simulation model

There are two important phases for a simulation model:

(a) Data generation: Data generation involves the sample observations of variables and can be carried out with the help of any of the following methods.

(i) Using the random number tables.

(ii) Restoring the mechanical devices.

(iii) Using the electronic computers.

(b) Book-keeping: This phase of simulation model deals with updating the system when new events occur, monitoring and recording the system states as and when they change, and keeping track of quantities of our interest to compute the measures of effectiveness.

Exercise 1: Consider a situation where customers arrive at a one-man barber shop for hair-cutting with the inter-arrival time of 15 minutes. In case the facility is not free, the customer has to wait till the facility becomes free. Customers are aware of the fact that for one cutting, the barber takes 20 minutes time and the shop opens from 8.00 Hrs to 12.00 Hrs. The customers will be asked to go back if they cannot be given service. Calculate the average waiting time for those customers who waited for the service. Also calculate the idle time for the service.

Solution: As the barber takes 20 minutes for one cutting, so only 12 customers will be entertained. The customers keep on arriving with the inter-arrival time of 15 minutes. Since the facility is available when the shop opens, the service of the first customer starts at 8.00 Hrs and the customer has not to wait. Arrival of customer 2 occurs at 8.15 Hrs. The service facility is still busy, the customer 2 waits and is the first one in the waiting queue. At 8.20 Hrs, the customer 1 departs, leaving the facility free. The service of customer 2 starts at 8.20 Hrs. Customer 3 arrives at 8.30 and waits. The service of customer 3 starts at 8.40 Hrs. The process of arrival and departure of customers continues till 12 customers arrive, as shown in the following table.

Simulation table

Time	Event	Customer	Time	Event	Customer
8.00	Arrival	1	9.45	Arrival	8
8.15	Arrival	2	10.00	Arrival	9
8.20	Departure	1	10.00	Departure	6
8.30	Arrival	3	10.15	Arrival	10
8.40	Departure	2	10.20	Departure	7
8.45	Arrival	4	10.30	Arrival	11
9.00	Arrival	5	10.40	Departure	8
9.00	Departure	3	10.45	Arrival	12
9.15	Arrival	6	11.00	Departure	9
9.20	Departure	4	11.20	Departure	10
9.30	Arrival	7	11.40	Departure	11
9.40	Departure	5	12.00	Departure	12

The waiting times for customers are as the following.

Waiting times

Customer	Waiting time (min)	Customer	Waiting time (Hrs)
2	8.20 − 8.15 = 5	8	10.20 − 9.45 = 35
3	8.40 − 8.30 = 10	9	10.40 − 10.00 = 40
4	9.00 − 8.45 = 15	10	11.00 − 10.15 = 45
5	9.20 − 9.00 = 20	11	11.20 − 10.30 = 50
6	9.40 − 9.15 = 25	12	11.40 − 10.45 = 55
7	10.00 − 9.30 = 30		Total = 330 min

Since 11 customers waited, the average waiting time is

$$\text{average waiting time} = \frac{330}{11} = 30$$

Thus, the average waiting time is 30 min per person. Since the service was always engaged, the idle time is zero.

Exercise 2: Consider a situation where customers arrive at a one-man barber shop for hair-cutting with the inter-arrival time of 15 minutes. In case the facility is not free, the customer has to wait till the facility becomes free. Customers are aware of the fact that for one cutting, the barber takes 20 minutes time and the shop opens from 8.00 Hrs and is closed at 12.00 Hrs all of the sudden. Calculate the average waiting time for those customers who waited for the service. Also calculate the idle time for the service.

Solution: As the customers know that the barber takes 20 minutes for one cutting, so no customer will arrive after 11.40 Hrs. On the other side, the customer do not how many are already in the waiting queue. The customers keep on arriving with the inter-arrival time of 15 minutes. Since the facility is available when the shop opens, the service of the first customer starts at 8.00 Hrs and the customer has not to wait. Arrival of customer 2 occurs at 8.15 Hrs. The service facility is still busy, the customer 2 waits and is the first one in the waiting queue. At 8.20 Hrs, the customer 1 departs, leaving the facility free. The service of customer 2 starts at 8.20 Hrs. Customer 3 arrives at 8.30 and waits. The service of customer 3 starts at 8.40 Hrs. The process of arrival and departure of customers continues up to 12.00 Hrs, as shown in the following table.

Simulation table

Time	Event	Customer	Time	Event	Customer
8.00	Arrival	1	10.00	Departure	6
8.15	Arrival	2	10.15	Arrival	10
8.20	Departure	1	10.20	Departure	7
8.30	Arrival	3	10.30	Arrival	11
8.40	Departure	2	10.40	Departure	8
8.45	Arrival	4	10.45	Arrival	12
9.00	Arrival	5	11.00	Arrival	13
9.00	Departure	3	11.00	Departure	9
9.15	Arrival	6	11.15	Arrival	14
9.20	Departure	4	11.20	Departure	10
9.30	Arrival	7	11.30	Arrival	15
9.40	Departure	5	11.40	Departure	11
9.45	Arrival	8	12.00	Departure	12
10.00	Arrival	9	12.00	Departure	13, 14, 15

It is obvious that the customers 13, 14 and 15 could not get the service, as the shop was closed suddenly. The waiting times for customers are as the following.

Waiting times

Customer	Waiting time (min)	Customer	Waiting time (min)
1	0.00	9	$10.40 - 10.00 = 40$
2	$8.20 - 8.15 = 5$	10	$11.00 - 10.15 = 45$
3	$8.40 - 8.30 = 10$	11	$11.20 - 10.30 = 50$
4	$9.00 - 8.45 = 15$	12	$11.40 - 10.45 = 55$
5	$9.20 - 9.00 = 20$	13	$12.00 - 11.00 = 60$
6	$9.40 - 9.15 = 25$	14	$12.00 - 11.15 = 45$
7	$10.00 - 9.30 = 30$	15	$12.00 - 11.30 = 30$
8	$10.20 - 9.45 = 35$		Total = 465 min

Since 14 customers waited, the average waiting time is

$$\text{average waiting time} = \frac{465}{14} = 33.21$$

Thus, the average waiting time is 33.21 min per person. Since the service was always engaged, the idle time is zero.

Exercise 3: Dr. Alexander is dentist and schedules per patient for 30 minutes appointment. Some of the persons may take more time and the others may take less time as compared to 30 minutes depending on the type of the work to be done. The following summary shows various categories of works, their probabilities and the time actually needed to complete the work.

Work =	Filling	Crown	Cleaning	Extraction	Check up
Time (min)	45	60	15	45	15
Probability	0.35	0.10	0.20	0.10	0.25

Simulate the dentist's clinic for the customers who arrive between 8.00 a.m. and 11.30 a.m. and determine the average waiting time for the patients and idle time of the dentist. The dentist offers service to all the patients who arrive within the stipulated time. Assume that all the patients reach the clinic at exactly their schedule arrival time starting at 8.00 a.m. Use the following random numbers for handling the simulation.

36 80 13 23 60 96 17 72

Solution: As the patients arrive with the interval of 30 min, during the period from 8.00 a.m. to 11.30 a.m., there will be 8 patients. Suppose the numbers $0 - 99$ are alloted in proportion to the probabilities associated with each category of work, then various kinds of work can be sampled, using random number table.

Work =	Filling	Crown	Cleaning	Extraction	Check up
Probability	0.35	0.10	0.20	0.10	0.25
Cum. Prob.	0.35	0.45	0.65	0.75	1.00
Rand No.	$0 - 34$	$35 - 44$	$45 - 64$	$65 - 74$	$75 - 99$
Total Nos.	35	10	20	10	25

Using the given random numbers, we can prepare a work-sheet as the following.

Patient	Arrival Time	Random No.	Work	Service Time (min)
1	8.00 a.m.	36	Crown	60
2	8.30 a.m.	80	Check up	15
3	9.00 a.m.	13	Filling	45
4	9.30 a.m.	23	Filling	45
5	10.00 a.m.	60	Cleaning	15
6	10.30 a.m.	96	Check up	15
7	11.00 a.m.	17	Filling	45
8	11.30 a.m.	72	Extraction	45

Now, we can simulate the dentist's clinic for the said period starting from 8.00 a.m. as the following.

Time	Event (Arr/Dep)	Patient No. (Service Time)	Patient Waiting
8.00 a.m.	1st Arr	1st (60 min)	-
8.30 a.m.	2nd Arr	1st (30 min)	2nd
9.00 a.m.	1st Dep, 3rd Arr	2nd (15 min)	3rd
9.15 a.m.	2nd Dep	3rd (45 min)	-
9.30 a.m.	4th Arr	3rd (30 min)	4th
10.00 a.m.	3rd Dep, 5th Arr	4th (45 min)	5th
10.30 a.m.	6th Arr	4th (15 min)	5th & 6th
10.45 a.m.	4th Dep	5th (15 min)	6th
11.00 a.m.	5th Dep, 7th Arr	6th (15 min)	7th
11.15 a.m.	6th Dep	7th (45 min)	-
11.30 a.m.	8th Arr	7th (30 min)	8th
12.00 noon	7th Dep	8th (45 min)	-
12.45 p.m.	8th Dep	-	-

This table shows that the dentist was not idle during the entire simulated period. The waiting times for the patients have been as the following.

Waiting times

Customer	Waiting time (min)
2	$9.00 - 8.30 = 30$
3	$9.15 - 9.00 = 15$
4	$10.00 - 9.30 = 30$
5	$10.45 - 10.00 = 45$

Customer	Waiting time (Hrs)
6	$11.00 - 10.30 = 30$
7	$11.15 - 11.00 = 15$
8	$12.00 - 11.30 = 30$
	Total $= 195$ min

Average waiting time per patient $= 195/7 = 27.86$ min.

Exercise 4: The rain fall in a city on a day depends upon whether there was rain fall on the previous day or not. When there was rain fall on the previous day, the probability of rain fall is as the following.

Rain fall =	No rain	1 cm	2 cm	3 cm	4 cm	5cm
Probability	0.40	0.25	0.20	0.10	0.03	0.02

When there was no rain fall on the previous day, the probability of rain fall is as the following.

Rain fall =	No rain	1 cm	2 cm	3 cm	4 cm
Probability	0.60	0.20	0.10	0.06	0.04

Simulate the city's whether for 10 days (assuming that there was no rain fall before the simulation period) determine the number of days without rain fall and the total rain fall during the period. Use the following random numbers for simulation.

56 63 38 55 65 78 88 08 78 96

Solution: We simulate the city's weather with and without rain fall in the following manner. The distribution when there was rain fall on the previous day is as the following.

Rain	Probability	Cum. Prob.	Rand No.	Total Nos.
No Rain	0.40	0.40	00 - 39	40
1 cm	0.25	0.65	40 - 64	25
2 cm	0.20	0.85	65 - 84	20
3 cm	0.10	0.95	85 - 94	10
4 cm	0.03	0.98	95 - 97	3
5 cm	0.02	1.00	98 - 99	2

The distribution when there was no rain fall on the previous day is as the following.

Rain	Probability	Cum. Prob.	Rand No.	Total Nos.
No Rain	0.60	0.60	00 - 59	60
1 cm	0.20	0.80	60 - 79	20
2 cm	0.10	0.90	80 - 89	10
3 cm	0.06	0.96	90 - 95	6
4 cm	0.04	1.00	96 - 99	4

On using the given random numbers, we can simulate for 10 days as the following.

Day	Rand No.	Rain	Cum. Rain
1	56	No Rain	-
2	63	1 cm	1 cm
3	38	No Rain	1 cm
4	55	No Rain	1 cm
5	65	1 cm	2 cm
6	78	2 cm	4 cm
7	88	3 cm	7 cm
8	08	No Rain	7 cm
9	78	1 cm	8 cm
10	96	4 cm	12 cm

During the simulation period, there was no rain on 4 days out of 10 days. The total rain fall during the period is 12 cm.

Exercise 5: The rain fall in a city on a day depends upon whether there was rain fall on the previous day or not. When there was rain fall on the previous day, the probability of rain fall is as the following.

Rain fall =	No rain	1 cm	2 cm	3 cm	4 cm	5cm
Probability	0.45	0.25	0.18	0.07	0.03	0.02

When there was no rain fall on the previous day, the probability of rain fall is as the following.

Rain fall =	No rain	1 cm	2 cm	3 cm	4 cm
Probability	0.60	0.20	0.10	0.06	0.04

Simulate the city's whether for 10 days, assuming that there was rain fall before the first day of the simulation period, and determine the number of days without rain and the total rain fall during the period. Use the following random numbers for simulation.

65 86 37 65 82 96 70 08 81 99

Solution: We simulate the city's weather with and without rain fall in the following manner. The distribution when there was rain fall on the previous day is as the following.

Rain	Probability	Cum. Prob.	Rand No.	Total Nos.
No Rain	0.45	0.45	00 - 44	45
1 cm	0.25	0.70	45 - 69	25
2 cm	0.18	0.88	70 - 87	18
3 cm	0.07	0.95	88 - 94	7
4 cm	0.03	0.98	95 - 97	3
5 cm	0.02	1.00	98 - 99	2

The distribution when there was no rain fall on the previous day is as the following.

Rain	Probability	Cum. Prob.	Rand No.	Total Nos.
No Rain	0.60	0.60	00 - 59	60
1 cm	0.20	0.80	60 - 79	20
2 cm	0.10	0.90	80 - 89	10
3 cm	0.06	0.96	90 - 95	6
4 cm	0.04	1.00	96 - 99	4

On using the given random numbers, we can simulate for 10 days as the following.

Day	Rand No.	Rain	Cum. Rain
1	65	1 cm	1 cm
2	86	2 cm	3 cm
3	37	No Rain	3 cm
4	65	1 cm	4 cm
5	82	2 cm	6 cm
6	96	4 cm	10 cm
7	70	2 cm	12 cm
8	08	No Rain	12 cm
9	81	2 cm	14 cm
10	99	5 cm	19 cm

During the simulation period, there was no rain on 2 days out of 10 days. The total rain fall during the period is 19 cm.

Exercise 6: A automobile company manufactures about 140 bikes per day. The daily production varies from 136 to 144 depending on the availability of raw materials and other working conditions.

Production =	136	137	138	139	140	141	142	143	144
Probability	.03	.09	.12	.14	.20	.15	.13	.10	.04

The finished bikes are transported in a specially arranged lorry which can accommodate 140 bikes. Using the following random numbers

70 92 65 44 53 41 08 17 02 84 07 95 88 75 47

simulate the process of transportation to find out: (i) what will be the average number of scooters waiting in the factory? (ii) What will be the average number of empty space in the lorry?

Solution: The random numbers are established as in the following table.

Production	Probability	Cum. Prob.	Rand No.	Total Nos.
136	0.03	0.03	00 - 02	3
137	0.09	0.12	03 - 11	9
138	0.12	0.24	12 - 23	12
139	0.14	0.38	24 - 37	14
140	0.20	0.58	38 - 57	20
141	0.15	0.73	58 - 72	15
142	0.13	0.86	73 - 85	13
143	0.10	0.96	86 - 95	10
144	0.04	1.00	96 - 99	4

Based on the given 15 random numbers, we simulate the process of production and transportation per day in the following manner.

S.No.	Rand No.	Production	Bikes waiting	Empty space
1	70	141	1	
2	92	143	3	
3	65	141	1	
4	44	140		
5	53	140		
6	41	139		1
7	8	137		3
8	17	138		2
9	2	136		4
10	84	142	2	
11	7	137		3
12	95	143	3	
13	88	143	3	
14	75	142	2	
15	47	140		
	Total		15	13

(i) Average number of bikes waiting $= 15/15 = 1.0$ per day

(ii) Average number of empty space $= 13/15 = 0.87$ per day

Exercise 7: A company has a single service station where the time between successive arrivals varies from one minute to six minutes and the service time varies from one minute to five minutes. The probabilities of arrival-interval and service time are as the following.

Arrival-interval (min)	Probability	Service time (min)	Probability
1	0.05	1	0.10
2	0.20	2	0.20
3	0.35	3	0.40
4	0.25	4	0.20
5	0.10	5	0.10
6	0.05		

The company starts at 11.00 a.m. and closes at 12.00 noon. A customer moves immediately into the service facility when it is free. On the other hand, if the service station is busy, the customer will wait. Calculate (i) average number of customers in the waiting queue, (ii) average waiting time, and (iii) average service time. Use the Monte-Carlo simulation technique with the random numbers

65	04	03	69	05	62	15	09	37	40	08
10	60	54	06	63	38	29	98	85		

for the arrival-interval and the random numbers

29	76	39	23	58	08	11	71	85	24	10
78	60	02	53	98	91	85	12	68		

for the service time.

Solution: From the given probabilities for arrival-interval and for service time, we first calculate the cumulative probabilities and the random numbers to be assigned as in the following table. These are then used as the basis for generating arrival and service times in conjunction with the given random numbers.

Arr.-int. (min)	Prob.	Cum. prob.	Rand Nos.	Service (min)	Prob.	Cum. prob.	Rand Nos.
1	0.05	0.05	00 - 04	1	0.10	0.10	00 - 09
2	0.20	0.25	05 - 24	2	0.20	0.30	10 - 29
3	0.35	0.60	25 - 59	3	0.40	0.70	30 - 69
4	0.25	0.85	60 - 84	4	0.20	0.90	70 - 89
5	0.10	0.95	85 - 94	5	0.10	1.00	90 - 99
6	0.05	1.00	95 - 99				

A simulation work-sheet is then prepared in the following manner.

Arrival			Service		Service	Waiting	No. of
RN	interval min	Time min	RN	duration min	start Hrs	time min	customers waiting
65	4	11.04	29	2	11.04	-	-
04	1	11.05	76	4	11.06	1	1
03	1	11.06	39	3	11.10	4	1
69	4	11.10	23	2	11.13	3	1
05	2	11.12	58	3	11.15	3	1
62	4	11.16	08	1	11.18	2	1
15	2	11.18	11	2	11.19	1	1
09	2	11.20	71	4	11.21	1	1
37	3	11.23	85	4	11.25	2	1
40	3	11.26	24	2	11.29	3	1
08	2	11.28	10	2	11.31	3	1
10	2	11.30	78	4	11.33	3	1
60	4	11.34	60	3	11.37	3	1
54	3	11.37	02	1	11.40	3	1
06	2	11.39	53	3	11.41	2	1
63	4	11.43	98	5	11.44	1	1
38	3	11.46	91	5	11.49	3	1
29	3	11.49	85	4	11.54	5	1
98	6	11.55	12	2	11.58	3	1
85	5	12.00	68	3	12.00	-	-
				59		46	18

$$\text{Average queue length} = \frac{18}{20} = 0.9$$

$$\text{Average waiting time} = \frac{46}{20} = 2.3 \text{ min}$$

$$\text{Average service time} = \frac{59}{20} = 2.95 \text{ min}$$

5. Random numbers

Random numbers (also known as random variates) are essentially the numbers distributed over the given interval in a random manner. They are not, however, truly random as they are generated systematically by an arithmetic process. One method of their generation is as the following. The random numbers in the interval [0, N] can be generated with the help of the following relation.

$$x_{n+1} = (rx_n + \mu)(\text{mod}N)$$

where x_n is the n-th random number, r a multiplier, and μ another positive integer (generally taken equal to zero).

In early days, manual methods were in use to generate the random numbers. The manual method can be used when the values of N and r are small. Besides the selection of N and r, here we have to assign the beginning of the random numbers, *i.e.,* x_1.

The procedure of the generation of random numbers can be understood with the help of the following exercises.

Exercise 8: Using the following relation, find out the random numbers where the starting number is 1, multiplier 13, additive number 0, and the highest value of a random number should not exceed the value 100.

$$x_{n+1} = (rx_n + \mu)(\text{mod}N)$$

Solution: Here, $x_1 = 1$, $r = 13$, $N = 100$ and we take $\mu = 0$. Other random numbers can be obtained with the help of the given relation as

$$x_2 = (13 \times 1 + 0)(\text{mod}100) = 13$$

$$x_3 = (13 \times 13 + 0)(\text{mod}100) = 69$$

$$x_4 = (13 \times 69 + 0)(\text{mod}100) = 97$$

$$x_5 = (13 \times 97 + 0)(\text{mod}100) = 61$$

$$x_6 = (13 \times 61 + 0)(\text{mod}100) = 93$$

$$x_7 = (13 \times 93 + 0)(\text{mod}100) = 9$$

$$x_8 = (13 \times 9 + 0)(\text{mod}100) = 17$$

$$x_9 = (13 \times 17 + 0)(\text{mod}100) = 21$$

$$x_{10} = (13 \times 21 + 0)(\text{mod}100) = 73$$

$$x_{11} = (13 \times 73 + 0)(\text{mod}100) = 49$$

$$x_{12} = (13 \times 49 + 0)(\text{mod}100) = 37$$

$$x_{13} = (13 \times 37 + 0)(\text{mod}100) = 81$$

$$x_{14} = (13 \times 81 + 0)(\text{mod}100) = 53$$

$$x_{15} = (13 \times 53 + 0)(\text{mod}100) = 89$$

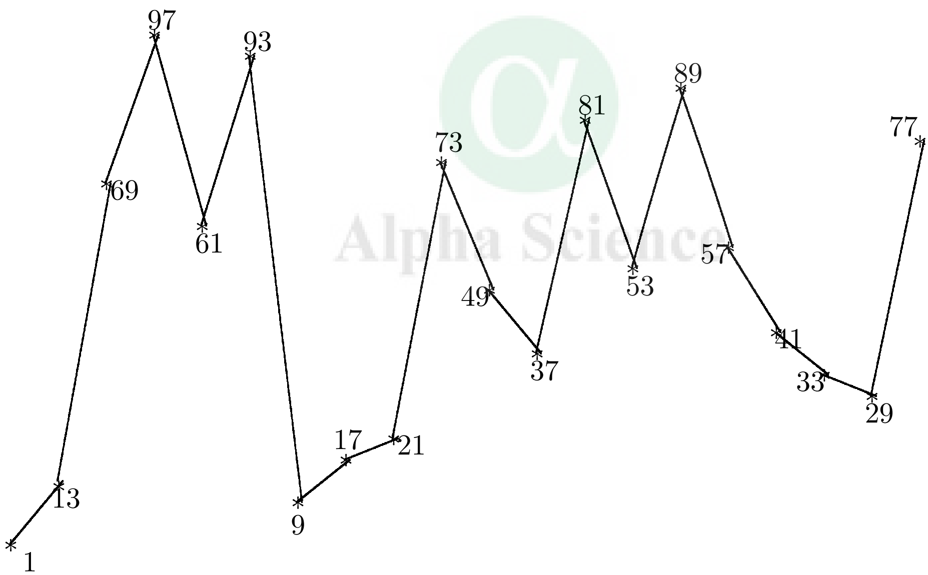
$$x_{16} = (13 \times 89 + 0)(\text{mod}100) = 57$$

$$x_{17} = (13 \times 57 + 0)(\text{mod}100) = 41$$
$$x_{18} = (13 \times 41 + 0)(\text{mod}100) = 33$$
$$x_{19} = (13 \times 33 + 0)(\text{mod}100) = 29$$
$$x_{20} = (13 \times 29 + 0)(\text{mod}100) = 77$$
$$x_{21} = (13 \times 77 + 0)(\text{mod}100) = 1$$

Hence, the random numbers generated are as the following.

1, 13, 69, 97, 61, 93, 9, 17, 21, 73, 49, 37, 81, 53, 89, 57, 41, 33, 29, 77, 1

There are 20 random numbers in the sequence which appear in a random manner. The sequence is however periodic in nature. The variation of these random numbers is shown in the following figure. The variation appears as random, and therefore, such numbers are known as random numbers.



Exercise 9: Using the following relation, find out the random numbers where the starting number is 1, multiplier 11, additive number 3, and the highest value of a random number should not exceed the value 110.

$$x_{n+1} = (rx_n + \mu)(\text{mod}N)$$

Solution: Here, $x_1 = 1$, $r = 11$, $N = 110$ and we take $\mu = 3$. Other random numbers can be obtained with the help of the given relation, as

$x_2 = (11 \times 1 + 3)(\text{mod}110) = 14$

$x_3 = (11 \times 14 + 3)(\text{mod}110) = 47$

$x_4 = (11 \times 47 + 3)(\text{mod}110) = 80$

$x_5 = (11 \times 80 + 3)(\text{mod}110) = 3$

$x_6 = (11 \times 3 + 3)(\text{mod}110) = 36$

$x_7 = (11 \times 36 + 3)(\text{mod}110) = 69$

$x_8 = (11 \times 69 + 3)(\text{mod}110) = 102$

$x_9 = (11 \times 102 + 3)(\text{mod}110) = 25$

$x_{10} = (11 \times 25 + 3)(\text{mod}110) = 58$

$x_{11} = (11 \times 58 + 3)(\text{mod}110) = 91$

$x_{12} = (11 \times 91 + 3)(\text{mod}110) = 14$

Hence, the random numbers generated are as the following.

1, 14, 47, 80, 3, 36, 69, 102, 25, 58, 91, 14

There are 11 random numbers in the sequence which appear in a random manner. The sequence is however periodic in nature, but the cycle excludes 1. It is interesting point.

Exercise 10: Using the following relation, find out the random numbers where the starting number is 1, multiplier 13, additive number 2, and the highest value of a random number should not exceed the value 100.

$$x_{n+1} = (rx_n + \mu)(\text{mod}N)$$

Solution: Here, $x_1 = 1$, $r = 13$, $N = 100$ and $\mu = 2$. Other random numbers can be obtained with the help of the given relation as

$x_2 = (13 \times 1 + 2)(\text{mod}100) = 15$

$x_3 = (13 \times 15 + 2)(\text{mod}100) = 97$

$x_4 = (13 \times 97 + 2)(\text{mod}100) = 63$

$x_5 = (13 \times 63 + 2)(\text{mod}100) = 21$

$x_6 = (13 \times 21 + 2)(\text{mod}100) = 75$

$x_7 = (13 \times 75 + 2)(\text{mod}100) = 77$

$x_8 = (13 \times 77 + 2)(\text{mod}100) = 3$

$x_9 = (13 \times 3 + 2)(\text{mod}100) = 41$

$$x_{10} = (13 \times 41 + 2)(\text{mod}100) = 35$$

$$x_{11} = (13 \times 35 + 2)(\text{mod}100) = 57$$

$$x_{12} = (13 \times 57 + 2)(\text{mod}100) = 43$$

$$x_{13} = (13 \times 43 + 2)(\text{mod}100) = 61$$

$$x_{14} = (13 \times 61 + 2)(\text{mod}100) = 95$$

$$x_{15} = (13 \times 95 + 2)(\text{mod}100) = 37$$

$$x_{16} = (13 \times 37 + 2)(\text{mod}100) = 83$$

$$x_{17} = (13 \times 83 + 2)(\text{mod}100) = 81$$

$$x_{18} = (13 \times 81 + 2)(\text{mod}100) = 55$$

$$x_{19} = (13 \times 55 + 2)(\text{mod}100) = 17$$

$$x_{20} = (13 \times 17 + 2)(\text{mod}100) = 23$$

$$x_{21} = (13 \times 23 + 2)(\text{mod}100) = 1$$

Hence, the random numbers generated are as the following.

1, 15, 97, 63, 21, 75, 77, 3, 41, 35, 57, 43, 61, 95, 37, 83, 81, 55, 17, 23, 1

There are 20 random numbers in the sequence which appear in a random manner. The sequence is however periodic in nature.

Exercise 11: Find out the random numbers, starting from a four digit number 1234, squaring the number and extracting four digits at the positions from third to sixth from right, *i.e.,* by using the mid-square method.

Solution: The random number and its square are given in Table 1.

Table 1

No.	number	square	No.	number	square
1	1234	1522756	29	8402	70593604
2	5227	27321529	30	5936	35236096
3	3215	10336225	31	2360	5569600
4	3362	11303044	32	5696	32444416
5	3030	9180900	33	4444	19749136
6	1809	3272481	34	7491	56115081
7	2724	7420176	35	1150	1322500
8	4201	17648401	36	3225	10400625
9	6484	42042256	37	4006	16048036
10	422	178084	38	480	230400
11	1780	3168400	39	2304	5308416
12	1684	2835856	40	3084	9511056
13	8358	69856164	41	5110	26112100
14	8561	73290721	42	1121	1256641
15	2907	8450649	43	2566	6584356
16	4506	20304036	44	5843	34140649
17	3040	9241600	45	1406	1976836
18	2416	5837056	46	9768	95413824
19	8370	70056900	47	4138	17123044
20	569	323761	48	1230	1512900
21	3237	10478169	49	5129	26306641
22	4781	22857961	50	3066	9400356
23	8579	73599241	51	4003	16024009
24	5992	35904064	52	240	57600
25	9040	81721600	53	576	331776
26	7216	52070656	54	3317	11002489
27	706	498436	55	24	576
28	4984	24840256	56	5	25

There are 56 random numbers

6. Problems and questions

1. Discuss about the need of simulation.

2. Consider a situation where customers arrive at a one-man barber shop for hair-cutting with the inter-arrival time of 15 minutes. However, no customer comes for one hour from 1.00 p.m. on ward. In case the facility is not free, the customer has to wait till the facility becomes free. Customers are aware of the fact that for

one cutting, the barber takes 25 minutes time and the shop opens from 8.00 a.m. to 6.00 p.m. Calculate the average waiting time for those customers who waited for the service. Also calculate the idle time for the service.

3. Using the following relation, find out the random numbers where the starting number is 1, multiplier 13, additive number 0, and the highest value of a random number should not exceed the value 100.

$$x_{n+1} = (rx_n + \mu)(\text{mod}N)$$

4. Find out the random numbers, starting from a four digit number 1234, squaring the number and extracting four digits at the positions from third to sixth from right, *i.e.,* by using the mid-square method.

5. Write short notes on the following:

   (i) Types of simulations

   (ii) Classification of simulation models

   (iii) Limitation of simulation technique

   (iv) Phases of simulation model

   (v) Random numbers

X. Computer in Research

To find out solution of a problem numerically, numerical calculations are carried out. Such task of numerical calculations has been made easy after the advent of computers. Most of the problems which could not be solved in absence of computers because of their large sizes, are being now solved with the help of computers, available nowadays. Present computers are quite efficient. In the present time, computers are being used in almost every walk of life. One cannot imagine research work in any field where computer is not essentially required. Right from the survey of literature, electronic communication, arrangement of data, calculations, analysis of results, writing the reports and finally preparation of thesis, and most of the other works are generally performed with the help of computers. Use of computers, along with the reliable and efficient softwares has made the task in each and every field very easy and comfortable. Presentations of lectures and seminars have become very effective. Thus, a computer is an integral part of research work as well. In view of this all, the present chapter is devoted for the introduction of basics of computer.

1. Computer generations

As the name indicates, a computer is nothing but a device which performs computation. In this sense, any device, however crude or sophisticated, which is used for mathematical computations becomes a computer. But what has made the term 'computer' conspicuous today and, what we normally imply when we speak of computers, are electronically operating machines which are used for doing calculations and manipulation of numbers. Thus, in brief, computer is an electronic device which is capable of receiving, storing, manipulating and yielding information such as numbers, words, pictures, etc.

Computers may be classified broadly into two categories: (i) digital computer, and (ii) analog computer. A digital computer is one which operates essentially by counting (using information, including letters and symbols, in the coded form) whereas the analog computer operates by measuring rather than counting. Digital computers handle information in the form of strings of binary numbers, *i.e.,* zeros and ones, with the help of counting process. On the other side, an analog computer converts the varying quantities, such as temperature and pressure into corresponding electrical voltages and then preforms specified functions on the given signals. The analog computers are generally used for certain specified engineering and scientific applications. Most of the computers we use nowadays are of digital type.

Computer technology has undergone tremendous modifications over a period of six decades. Nowadays, a microcomputer is far more powerful and cost very little as compared to the world's first electronic computer, *viz.,* Electronic Numerical Integrator and Calculator (ENIAC) completed in 1946. A personal computer (PC) works many times faster and is much more reliable and has very large memory as compared to ENIAC.

The developments in the computer technology are usually talked in terms of generations.[1] Nowadays, we have fourth generation computers in service and further efforts are being made to develop the fifth generation computers. The first generation computer, started in 1945, contained 18000 small bottle-sized valves which constituted its central processing unit (CPU). That machine did not have any provision for storing programs and instructions. With the invention of transistors in 1947, in the second generation computers, the valves were replaced by the transistors. With this, the size of a computer reduced a lot and made it more reliable. These computers appeared in the market in early 1960's. The third generation computers followed the invention of integrated circuit (IC) in 1959. Such computers with their CPU and main store made up of IC chips, were available in the market in the second half of 1960's. The fourth generation computers are due to the advent of microprocessor in 1972. The use of microprocessor as a CPU in a computer has made real the dream of 'computer for the masses'. This device has enabled the development of microcomputers, personal computers, portable computers, and so on. The fifth generation is expected to be 50 times or so more faster than the present day super-fast computer.

In the present day computers, the input is through the Visual Display

[1]First generation computers were manufactured between 1945-60, such as IBM 650, IBM 701. Second generation computers were manufactured between 1960-65, such as IBM 1401, Honeywell 40. Third generation computers were manufactured between 1965-70, such as IBM System 360, 370. Fourth generation computers are those which manufactured between 1971 to this date, such as IBM 3033, HP 3000 and many more.

Unit (VDU) which consists of a TV-like screen and a key-board which is used for feeding data into a computer. The output can be printed on the low-cost high-speed printers. VDU may also be used as an output device. The data can be stored on bubble memories and optical video discs. In brief, computer technology has become highly sophisticated and is being developed further at a very fast speed.

2. Computer system

Computer is an electronic device which can perform a variety of tasks assigned to it. Computer does not perform any thing from its own, but it follows sincerely the instructions supplied to it. Hence, a computer is often referred to as an obedient servant. As one can see, a computer can be broadly considered as having four main components (Figure 1): (i) Input unit, (ii) Central processing unit, (iii) Memory unit and (iv) Output unit.

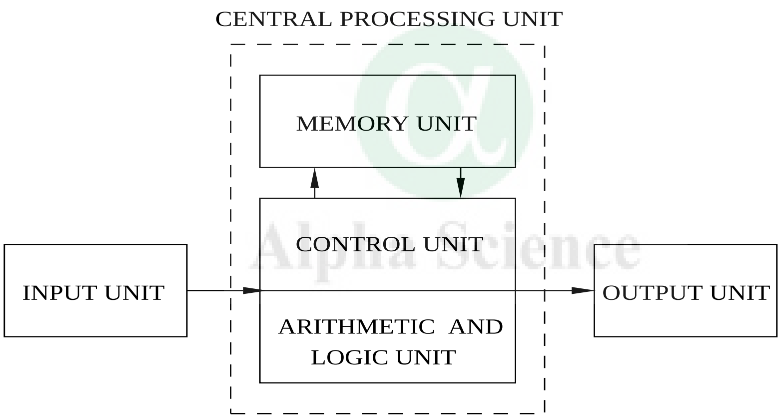


Figure 1: Block diagram of a computer.

2.1 Input unit

The input unit allows the user to enter data as well as instructions to the main part of the computer. The input unit therefore acts as an interface between the user and the computer. A variety of input devices are in use. Some commonly used input devices are the keyboard, mouse, light-pen, floppy disk drive, CD drive, etc.

2.2 Central processing unit

The main part of the computer is known as the central processing unit (CPU). It processes the data supplied to it as per instructions (program) and keeps the control over the functioning of all the peripherals (input/output devices). The CPU comprises mainly of two units: (i) Control unit and (ii) Arithmetic and logic unit (ALU). The CPU can be considered like the mind in a human body.

(i) Control unit

The control unit makes coordination among different parts of a computer. It collects instructions and data from the input unit as well as the memory and then generates a sequence of timing and control signals. These signals direct the input, output, ALU as well as the memory units to carry out the desired operations.

(ii) Arithmetic and logic unit

The arithmetic and logic unit (ALU) performs arithmetic operations such as addition, subtraction, multiplication, etc. as well as performs the relational and logical operations. For scientific calculations, the math co-processor is employed in the CPU.

2.3 Memory unit

The memory unit provides space for storing data as well as instructions. Before their processing, the input data are also stored temporarily in the memory unit. During the processing, the intermediate results as well as the final results, before sending them to the output unit, are also stored in the memory unit.

The memory may be classified into two categories: (i) primary memory and (ii) secondary memory. The primary memory is directly accessible to the CPU. Often the primary memory is considered as a part of the CPU. The secondary memory is an additional memory used for the storage of data and programs. Some examples of the secondary storage devices are the CD, DVD, pen-drive etc.

2.4 Output unit

After processing the input data by the CPU, the results obtained are sent to the output unit where the data are either stored in the secondary storage devices or printed on a paper. The results may also be displayed on the console (Video Display Unit).

3. Coding system

A computer program (set of instructions) is written in a language, like FORTRAN, BASIC or C, where roman alphabets, decimal digits and specific characters are used. Each of these language is known as the High Level Language (HLL). But, a computer cannot understand these characters (roman alphabets, decimal digits and specific characters). A computer can however distinguish between the two situations, which can be understood with the help of an example giving two different situations, such as a lamp is ON or OFF; one specific pole of a magnet (north or south pole) is UP or DOWN. Such two situations, in the mathematical language are denoted by 0 and 1. Thus, one can say that the computer understands a binary system whose basic digits are 0 and 1. In order to make an equivalence between the two situations, all the characters are therefore coded in terms of the binary digits. There are two main coding systems:

(i) American Standard Code for Information Interchange (ASCII)

(ii) Extended Binary Coded Decimal Information Code (EBCDIC)

The coding in the ASCII code is as shown in Table 1. Now, we have a situation where we write a computer program by using various characters (roman alphabets, decimal digits and specific characters), but the computer understands the corresponding codes written in the binary system. This situation is very similar to that, for example, Mr. A visits Mr. B and they do not know any common language and therefore cannot speak directly with each other. However, they can speak to each other with the help of a translator who can translate efficiently the language of Mr. A into the language of Mr. B, and vice versa. For example, English language is translated into Russian language and vice versa.

In the context of a computer, the languages of Mr. A and Mr. B are the aforesaid characters and their binary codes. The work of a translator in a computer is performed by a compiler or an interpreter. A compiler and an interpreter are softwares which translate the characters into the binary code (called machine language) and vice versa.

Table 1. ASCII coding of the characters

Char	ASCII	Char	ASCII	Char	ASCII
A	0100 0001	g	0110 0111	#	0010 0011
B	0100 0010	h	0110 1000	$	0010 0100
C	0100 0011	i	0110 1001	%	0010 0101
D	0100 0100	j	0110 1010	&	0010 0110
E	0100 0101	k	0110 1011	′	0010 0111
F	0100 0110	l	0110 1100	(	0010 1000
G	0100 0111	m	0110 1101	)	0010 1001
H	0100 1000	n	0110 1110		0010 1010
I	0100 1001	o	0110 1111	+	0010 1011
J	0100 1010	p	0111 0000	,	0010 1100
K	0100 1011	q	0111 0001	-	0010 1101
L	0100 1100	r	0111 0010	.	0010 1110
M	0100 1101	s	0111 0011	/	0010 1111
N	0100 1110	t	0111 0100	:	0011 1010
O	0100 1111	u	0111 0101	;	0011 1011
P	0101 0000	v	0111 0110	<	0011 1100
Q	0101 0001	w	0111 0111	=	0011 1101
R	0101 0010	x	0111 1000	Blank	0010 0000
S	0101 0011	y	0111 1001	>	0011 1110
T	0101 0100	z	0111 1010	?	0011 1111
U	0101 0101	0	0011 0000	@	0100 0000
V	0101 0110	1	0011 0001	[	0101 1011
W	0101 0111	2	0011 0010	\	0101 1100
X	0101 1000	3	0011 0011	]	0101 1101
Y	0101 1001	4	0011 0100	∧	1001 1110
Z	0101 1010	5	0011 0101	_	0101 1111
a	0110 0001	6	0011 0110	‘	0110 0000
b	0110 0010	7	0011 0111	{	0111 1011
c	0110 0011	8	0011 1000	|	0111 1100
d	0110 0100	9	0011 1001	}	0111 1101
e	0110 0101	!	0010 0001		0111 1110
f	0110 0110	”	0010 0010	Delete	0111 1111

4. Translators

As discussed, in a computer, for the translation between the characters and their binary codes, we need a translator. Now, the translation can be done in two ways. To understand these two ways, suppose in the aforesaid example, Mr. A wants to deliver a talk and Mr. B wants to listen it with the help of a translator. In one way of translation of the speech, each sentence is translated before the next sentence is spoken. In another way, the translation is done at the end of the complete speech.

In the computer, the first way of the translation is done by an interpreter which translates a computer program piece-wise. The second way of translation is done by a compiler which translates the complete program as a whole at the end. The use of an interpreter or a compiler depends on the computer language. For example, in BASIC we use an interpreter whereas for FORTRAN and C we use a compiler.

4.1 Interpreter

Interpreter is a software program which translates a piece of the program written in a HLL at a time to an equivalent machine code; again it takes next piece of the program to translate it. This process continues till the complete program is translated. It means that the selection of a part of the program and then translation is carried out by the interpreter alternatively. As every time, a part of the program is translated into a machine code, an interpreter is sensitive to the fast changes in the source program. This is the main advantage of the use of an interpreter over a compiler. As compared to a compiler, an interpreter is less complicated and it requires very less memory space, but more time for translation and then execution of the program.

4.2 Compiler

Compiler is a rather complicated software program which translates a program written in a HLL as a whole to the equivalent machine code and then executes it. Separate compiler programs are written for each high level language as different symbols and syntax are used for a different language as well as for a different version of the language. The compiler program can be stored permanently on the secondary storage and is copied into the main memory of the computer for its use as per requirement. The compiler only compiles the given program and it is not involved in the execution of the program. While translating a given program it compiles the whole source program into equivalent object code. The object code is saved in the memory and is called for its use when the program is executed. As the whole program is compiled at one time, it requires more time and large memory space. The compiled program is executed in a rather less time. When there is any error in the source program the compiler points out through a suitable error message. After making the corrections of the error, the compilation is done as a fresh.

5. Some more terms

Let us discuss about some more terms in the context of computers as the following.

(a) **Hardware:** All the physical components, such as CPU, input devices, output devices, storage devices, etc. come under the category of hardware.

(b) **Software:** All the computer programs (sets of instructions) come under the category of software.

(c) **Firmware:** It is a software which is incorporated into the electronic circuitry of computer by the manufacturer.

(d) **System software:** It is a software which tell the computer how to perform a function. It is often known as the operating software and is normally supplied by the computer manufacturer.

(e) **Application software:** It is a software which tells the computer how to perform the specific tasks such as preparation of company pay roll, calculation of a mathematical expression, etc. These softwares are written by the user of the computer or supplied by the software companies which produce and sell the softwares for the complicated tasks.

(f) **Integrated circuit (IC):** It is a complete electronic circuit which has been fabricated on a single piece of pure silicon. Silicon is the most commonly used semiconductor.[2]

(g) **Memory chips:** The IC's form a secondary memory or storage of the computer. They retain data and instructions (programs) which are not needed immediately by the main memory, and is contained in the CPU.

(h) **Two-state devices:** The transistors of an IC chip can take only two states, which may be understood as either ON or OFF, either conducting or non-conducting. The ON state is conventionally represented by one (1) whereas the OFF state by zero (0). These

[2]Semiconductor is a material whose conductivity lies in between the conductivities of a conductor and an insulator. Thus, a semiconductor is neither a good conductor nor a bad conductor.

two numbers, 0 and 1, form a binary system and these two binary digits are known as the bits. Thus, 0 and 1 are two bits of the binary system. A string of eight bits is known as a byte and a group of byte constitutes a word. A chip is called, 8-bit, 16-bit, 32-bit and so on, depending on the number of bits combined in its standard word.

6. Characteristics of a computer

A computer has the following characteristics.

(a) Speed: Speed of present-day computer is so high that it can perform calculations in just a few seconds for which human being requires weeks time for doing by hand. This characteristic made many scientific projects feasible which were almost impossible in absence of computers.

(b) Diligence: Being a machine, a computer does not feel tired and does not show the lack of concentration. When millions of calculations are to be made, it will do all of them with the same accuracy, efficiency and speed. It is not the case that the first calculation is done with more accuracy, efficiency and speed as compared to the last one.

(c) Storage: Since the advent of the computers, the storage capacity of a computer has increased tremendously. However, it is still much smaller than that of a human brain. Hence, it is impossible to store all types of information inside the computer memory. Therefore, the information which is not presently required can be stored in an auxiliary storage device and the same may be brought into the main internal memory of the computer, as and when required for its use.

(d) Accuracy: The accuracy of a computer is very high. Error in a machine is probable, but due to increased efficiency in error-detecting techniques, a computer seldom produces false results. Errors in the calculations are generally due to human mistake rather than technological reasons. That is, the reason for wrong results is generally some errors in the computer program written by the user.

(e) Automation: Once a program is stored into the computer memory, the computer follows automatically all the instructions written in the program to the control unit for execution.

(f) Binary digits: Computers use the binary number system (a system in which all the characters are represented by the combinations of two digits - one (1) and zero (0). Thus, the base of the numbers used in the computers is two, whereas the ordinary decimal arithmetic system works on a base of ten.

7. Binary number system

The arithmetic system which uses two numbers (0 and 1), instead of ten numbers (0 - 9), but operates on the same logic is known as a binary system. Hence, the base of the binary system is 2 whereas the base of the decimal system is 10. The binary digits, 0 and 1, used in the system are known as the bits. The name ' binary' is given as it allows only two digits for the formation of numbers. Some examples of binary numbers are 101, 11010, 1001. Notice that only two digits 0 and 1 are used in the formation of binary numbers.

In the decimal system, the first place is for 1s, second place is for 10s and the third place is for 100s, and so on. For example, the number 427 in the decimal system can be expressed as

$$427 \text{ (decimal)} = 4 \times 10^2 + 2 \times 10^1 + 7 \times 10^0$$

In the binary system, we use 2 instead of 10, *i.e.,* the first place is still for 1s, but the second place is for 2s and the third place is for 4s, and so on. For example, the number 111 in the binary system can be expressed as

$$111 \text{ (binary)} = 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 7 \text{ (decimal)}$$

Thus, it also shows the equivalence between 111 (binary) and 7 (decimal). The procedures for conversion from the decimal system to binary system and vice versa are discussed in the following sections.

7.1 Decimal to binary conversion

A positive decimal integer can be easily converted into its equivalent binary number by repeated division by 2. The method is as the following. Divide the given positive integer by 2. Let R_1 be the remainder and q_1 the quotient. Remember that after division by 2, the remainder can either be zero (0) or one (1). Next divide q_1 by 2. Let R_2 be the remainder and q_2 the quotient. Continue this process of dividing the quotient by 2 until zero (0) is obtained as quotient. The equivalent binary number be formed by arranging the remainders as

$$R_k\ R_{k-1}\ \ldots R_2\ R_1$$

where R_k is the last remainder and R_1 the first remainder, obtained by division process. That is, the first remainder is the right-most digit of the binary number.

Exercise 1: Find the binary number equivalent of the decimal number 53.

Solution: For conversion of 53 (decimal) into the equivalent binary number, we prepare the following table for remainder and quotient.

Number to be divided by 2	Quotient	Remainder
53	26	1
26	13	0
13	6	1
6	3	0
3	1	1
1	0	1

On collecting the remainders obtained in the above table, we find

53 (decimal) = 110101 (binary)

Alternative method: We can write the given decimal number in terms of 1, 2, 4, 8, 16, 32, 64, 128, etc. For example, 53 can be expressed as the following.

$$\begin{aligned} 53 &= 32 + 16 + 0 + 4 + 0 + 1 \\ &= 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 110101 \text{ (binary)} \end{aligned}$$

7.2 Binary to decimal conversion

A simple method for conversion of a binary number to its decimal equivalent is as the following. Double the leftmost bit of the given binary number and add to it the bit at its right. Then again double the sum and add to it the bit at its right. Proceed in this manner till all the bits have been used. The final sum obtained by such repeated doubling and adding is the desired decimal equivalent.

Exercise 2: Find the decimal equivalent of 110101 (binary).

Solution: Doubling the leftmost digit we get 2.

Adding to it the bit on its right we get 2 + 1 = 3.

Doubling this number we get $3 \times 2 = 6$.

Adding to it the bit on the right we get $6 + 0 = 6$.

Doubling this number we get $6 \times 2 = 12$.

Adding to it the bit on the right we get $12 + 1 = 13$.

Doubling this number we get $13 \times 2 = 26$.

Adding to it the bit on the right we get $26 + 0 = 26$.

Doubling this number we get $26 \times 2 = 52$.

Adding to it the bit on the right we get $52 + 1 = 53$.

Thus, the binary number 110101 is equivalent to 53 in the decimal system.

Alternative method: We can write the given binary number as the following.

$$\begin{aligned} 110101 &= 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 32 + 16 + 0 + 4 + 0 + 1 = 53 \text{ (decimal)} \end{aligned}$$

Exercise 3: Convert the following binary numbers into their decimal equivalent.

(i) 1110, (ii) 11101, (iii) 11011, (iv) 111011

Solution: The decimal equivalent are obtained as the following.

$$1110 \text{ (binary)} = 1 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 14 \text{ (decimal)}$$

$$\begin{aligned} 11101 \text{ (binary)} &= 1 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 29 \text{ (decimal)} \end{aligned}$$

$$\begin{aligned} 11011 \text{ (binary)} &= 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\ &= 27 \text{ (decimal)} \end{aligned}$$

$$\begin{aligned} 111011 \text{ (binary)} &= 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\ &= 59 \text{ (decimal)} \end{aligned}$$

8. Calculations in the binary system

In a computer all the basic operations $(+, -, \times, \div)$ for the binary numbers are performed in the form of addition operation of binary numbers. For example, the multiplication 5×6 is performed in terms of addition as $6+6+6+6+6$. Subtraction and division operations are performed essentially by addition using the principle of complementation.

8.1 Addition of binary numbers

The basic rules for addition of binary numbers are as the following.

$$\begin{array}{r} 0 \\ +\ 0 \\ \hline 0 \\ \hline \end{array} \qquad \begin{array}{r} 0 \\ +\ 1 \\ \hline 1 \\ \hline \end{array} \qquad \begin{array}{r} 1 \\ +\ 0 \\ \hline 1 \\ \hline \end{array} \qquad \begin{array}{r} 1 \\ +\ 1 \\ \hline 1\ 0 \\ \hline \end{array}$$

Note that the sum of 1 and 1 is 10 (sum is 0 with carry 1) which is equivalent of the decimal digit 2.

Exercise 4: Add the binary numbers 1011 and 110. Show the equivalent to the addition in the decimal system.

Solution: Decimal equivalents of the binary numbers 1011 and 110 are respectively 11 and 6. The sum of the decimal numbers is $11 + 6 = 17$. The addition of binary numbers is as the following.

$$\begin{array}{r} 1011 \\ +\ 110 \\ \hline 10001 \\ \hline \end{array}$$

The binary number 10001 is equivalent to 17 in the decimal system.

Exercise 5: Add the binary numbers 1100110 and 1010111.

Solution: The addition of binary numbers 1100110 and 1010111 is as the following.

$$\begin{array}{r} 1100110 \\ +\ 1010111 \\ \hline 10111101 \\ \hline \end{array}$$

Here, for 1 + 1 + 1 we have 1 and carry 1.

Exercise 6: Perform the following additions.

(i) 1100 and 1110
(ii) 1111 and 1101
(iii) 1011 and 111
(iv) 1001 and 101

Solution: The additions of the binary numbers are as the following.

$$\begin{array}{r} 1100 \\ +\ 1110 \\ \hline 11010 \\ \hline \end{array} \qquad \begin{array}{r} 1111 \\ +\ 1101 \\ \hline 11100 \\ \hline \end{array} \qquad \begin{array}{r} 1011 \\ +\ 111 \\ \hline 10010 \\ \hline \end{array} \qquad \begin{array}{r} 1001 \\ +\ 101 \\ \hline 1110 \\ \hline \end{array}$$

8.2 Subtraction of binary numbers

Suppose the binary number B is to be subtracted from the binary number A. The subtraction of B from A is carried out in the following manner. This method is often known as the complementary subtraction method.

Step 1: Express both binary numbers A and B into the same number of digits. We can add zeros on the left side.

Step 2: Find the complement of B. Complement of a number is obtained by replacing 1 by 0 and 0 by 1. Let the complement of B be the binary number C.

Step 3: Add the binary numbers A and C.

Step 4: In the addition, if there is carry 1 in the final stage, add the carry 1 to the result to obtain the final result. When there is no carry 1, take complement of the result and put a negative sign to get the final result.

Exercise 7: Subtract 1011 from 11011. Show the equivalent subtraction in the decimal system.

Solution: Decimal equivalent of the binary numbers 1011 and 11011 are respectively 11 and 27. The subtraction of 11 from 27 gives 16. The subtraction of binary numbers is as the following. Both the numbers are made to have five digits, and thus we have 01011 and 11011, respectively. The complement of 01011 is 10100. Now, addition of 10100 and 11011 gives

$$\begin{array}{r} 10100 \\ +11011 \\ \hline 101111 \\ \hline \end{array}$$

The sum has six digits, showing that there is carry 1. On adding 1 to the five-digit result, we have

$$\begin{array}{r} 01111 \\ +1 \\ \hline 10000 \\ \hline \end{array}$$

This is the final result. The equivalence of 10000 is 16 in the decimal system.

Exercise 8: Subtract 11011 from 1011. Show the equivalent subtraction in the decimal system.

Solution: Decimal equivalent of the binary numbers 11011 and 1011 are respectively 27 and 11. The subtraction of 27 from 11 gives −16. The subtraction of binary numbers is as the following. Both the numbers are made to have five digits, and thus we have 11011 and 01011, respectively. The complement of 11011 is 00100. Now, addition of 00100 and 01011 gives

$$\begin{array}{r} 00100 \\ +01011 \\ \hline 01111 \\ \hline \end{array}$$

The sum has five digits, showing that there is no carry 1. Now, the complement of 01111 is 10000. On adding a negative sign, the final result is − 10000. The equivalent of − 10000 is − 16 in the decimal system.

Remark: The computer performs the division operation by repeating the complementary subtraction method. For example, $32 \div 8$ may be thought of as $32 - 8 = 24 - 8 = 16 - 8 = 8 - 8 = 0$ (*i.e.*, four times subtraction of 8).

8.3 Binary fraction

Just as we use a decimal point to separate the whole and fraction parts of a decimal number, we can use binary point in the binary numbers to separate the whole and fraction parts.

Conversion from binary to decimal

The binary fraction can be converted into the decimal fraction in the following manner.

$$\begin{aligned} 0.111 \text{ (binary)} &= (1 \times 2^{-1}) + (1 \times 2^{-2}) + (1 \times 2^{-3}) \\ &= 0.5 + 0.25 + 0.125 = 0.875 \text{ (decimal)} \end{aligned}$$

Conversion from decimal to binary

For conversion of decimal fraction into the binary fraction, we follow the following steps.

Step 1: Multiply the decimal fraction by 2 repeatedly. The whole number part of the first multiplication gives the first 1 or 0 of the binary fraction.

Step 2: The fractional part of the result is carried over and multiplied by 2. The whole number part of the multiplication gives the second 1 or 0 of the binary fraction.

Step 3: Step 2 is repeated until there is fractional part.

Step 4: Suppose the whole number parts received are w_1, w_2, ..., w_k, then the fractional part in binary system is

$$.w_1\ w_2\ \ldots w_k$$

Exercise 9: Convert 0.8125 into its equivalent binary fraction.

Solution: Multiply the decimal fraction by 2 and taking the whole number part repeatedly, we get as the following.

Number to be multiplied by 2	Value	Whole number	Fraction part
0.8125	1.625	1	0.625
0.625	1.25	1	0.25
0.25	0.50	0	0.50
0.50	1.0	1	0.00

Finally, the fraction part is zero. Thus, the equivalent binary fraction is

0.1101

Exercise 10: Convert decimal number 4.625 into its equivalent binary number.

Solution: The integer and fractional part of the decimal number is converted into its binary equivalents separately. decimal number 4 is equivalent to 100 binary number. Further, 0.625 decimal fraction is equivalent to 0.101 binary fraction. Thus, we have

4.625 (decimal) = 100.101 (binary)

9. Use of computers

It may be difficult to find any field where computer is not being used. It is being used in education, bank, industry, business, administration, transport, hospitals, social work, judiciary, and many more fields. Use of a computer made our life very comfortable and fast. Nowadays, for example, communication has become so fast that within a time of seconds, we can communicate with any person, having compatible communication facility, in any part of the world, and can get the required information. Here, we shall discuss about some applications where computer helps a researcher in his/her research work.

(a) Selection of institution:

For doing research work, one wants to have information about the institutions where the research work in the filed of ones interest is going on. Such information and other relevant details are available on the respective websites. Even if one does not know about the address of website of a particular institution, one can reach to it with the help of the 'google', for example. After reaching the homepage of a particular institution, one can find detailed information about the fields of research work going on their, about the faculty involved in the research work, fee structure and other charges etc. and so on.

(b) Review of literature:

In the old time, most of the information about the published research work was in the print form. Owing to the tremendous amount involved, every institution could not subscribe for complete set of literature. Therefore, for the review of literature, one had to take permission and visit various institutions. One had obviously to spend time and money, and to make plan and permission for the visit. That is, one could not find the required information as and when was required.

On the other side, nowadays, with the development of Internet facility, it is possible to access most of the information pertaining to research journals and related information. We can keep on reviewing day-to-day information about the research publications of various institutions. Sometimes scientists put their findings on the websites in order to get comments from the peer group.

(c) Correspondence:

In the old time, one had to write a letter and wait for the response. Now, with the help of Internet, scientists are communicating fast with one another. Even conferences are being organized while the scientists are sitting in various parts in the world. The research papers, which were being sent through mail in the past, are now being submitted through Internet; reports from the referees on the manuscripts of research papers are being invited through Internet and the response are being received through Internet. Use the use of Internet nowadays is so common and essential that scientists feel very odd when the Internet is not available even for a short time when required.

(d) Numerical calculations:

Numerical calculations have been the prime task for the computers right from the inception of computers. With the high speed and large memory size of a computer, it has now been possible to deal with very complicated models. In the computers, we can store a large amount of data, which can be retrieved as and when required. Scientific groups in various institutions (even situated in various countries) are collaborating and they often exchange data with the help of Internet.

(e) Preparation of manuscript:

Nowadays, preparation of manuscripts for research papers and thesis/dissertation has become very convenient. Very elegant softwares are available for preparing diagrams, graphs and some special effects. The manuscripts can be prepared in various colours and it makes a presentation very attractive and effective.

(f) Presentations:

For seminars, invited talks etc., the presentations are nowadays being prepared with the help of computers. In the presentations, we can included special effects as well.

10. Problems and questions

1. Write in brief about the computer generations.
2. Write in brief about the computer system.
3. Write in brief about the characteristics of a computer
4. Find out the binary equivalent for the decimal numbers 33, 57, 208
5. Find out the decimal equivalent for the binary numbers 1101, 10101
6. Find the sum of the binary numbers 11011 and 10011.

7. Subtract the binary number 1001 from 11011 and 11010 from 1011.

8. Write in brief about the computer applications

9. Find out the decimal equivalent for the binary fraction 0.1101

10. Find out the binary equivalent for the decimal fraction 0.125.

11. Write short note on the following:

    (i) Computer generations
    (ii) Computer system
    (iii) Input units
    (iv) Central processing unit
    (v) Memory unit
    (vi) Output units
    (vii) Translators
    (viii) Interpreter
    (ix) Compiler
    (x) Coding system
    (xi) Hardware and software
    (xii) Firmware, system software and application software
    (xiii) Integrated circuit and memory chips
    (xiv) Bits and byte

Appendices

Table 1A. Value of integral for different values of z_0

z_0	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	0.00
-3.0	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
-2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
-2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
-2.7	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
-2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047
-2.5	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062
-2.4	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082
-2.3	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107
-2.2	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139
-2.1	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179
-2.0	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228
-1.9	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287
-1.8	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359
-1.7	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446
-1.6	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548
-1.5	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668
-1.4	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808
-1.3	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968
-1.2	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151
-1.1	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357
-1.0	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587
-0.9	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841
-0.8	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119
-0.7	.2148	.2177	.2206	.2236	.2266	.2297	.2327	.2358	.2389	.2420
-0.6	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743
-0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085
-0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446
-0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821
-0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207
-0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602
0.0	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000

Table 1B. Value of integral for different values of z_0

z_0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986

Table 2. Percentile of the t-distribution

d.f.	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.3138	12.706	31.821	63.657
2	1.886	2.9200	4.3027	6.965	9.9248
3	1.638	2.3534	3.1825	4.541	5.8409
4	1.533	2.1318	2.7764	3.747	4.6041
5	1.476	2.0150	2.5706	3.365	4.0321
6	1.440	1.9432	2.4469	3.143	3.7074
7	1.415	1.8946	2.3646	2.998	3.4995
8	1.397	1.8595	2.3060	2.896	3.3554
9	1.383	1.8331	2.2622	2.821	3.2498
10	1.372	1.8125	2.2281	2.764	3.1693
11	1.363	1.7959	2.2010	2.718	3.1058
12	1.356	1.7823	2.1788	2.681	3.0545
13	1.350	1.7709	2.1604	2.650	3.0123
14	1.345	1.7613	2.1448	2.624	2.9768
15	1.341	1.7530	2.1315	2.602	2.9467
16	1.337	1.7459	2.1199	2.583	2.9208
17	1.333	1.7396	2.1098	2.567	2.8982
18	1.330	1.7341	2.1009	2.552	2.8784
19	1.328	1.7291	2.0930	2.539	2.8609
20	1.325	1.7247	2.0860	2.528	2.8453
21	1.323	1.7207	2.0796	2.518	2.8314
22	1.321	1.7171	2.0739	2.508	2.8188
23	1.319	1.7139	2.0687	2.500	2.8073
24	1.318	1.7109	2.0639	2.492	2.7969
25	1.316	1.7081	2.0595	2.485	2.7874
26	1.315	1.7056	2.0555	2.479	2.7787
27	1.314	1.7033	2.0518	2.473	2.7707
28	1.313	1.7011	2.0484	2.467	2.7633
29	1.311	1.6991	2.0452	2.462	2.7564
30	1.310	1.6973	2.0423	2.457	2.7500
35	1.3062	1.6896	2.0301	2.438	2.7239
40	1.3031	1.6839	2.0211	2.423	2.7045
45	1.3007	1.6794	2.0141	2.412	2.6896
50	1.2987	1.6759	2.0086	2.403	2.6778
60	1.2959	1.6707	2.0003	2.390	2.6603
70	1.2938	1.6669	1.9945	2.381	2.6480
80	1.2922	1.6641	1.9901	2.374	2.6388
90	1.2910	1.6620	1.9867	2.368	2.6316
100	1.2901	1.6602	1.9840	2.364	2.6260
120	1.2887	1.6577	1.9799	2.358	2.6175
140	1.2876	1.6558	1.9771	2.353	2.6114
160	1.2869	1.6545	1.9749	2.350	2.6070
180	1.2863	1.6534	1.9733	2.347	2.6035
200	1.2858	1.6525	1.9719	2.345	2.6006
∞	1.282	1.645	1.96	2.326	2.576

Table 3. Percentile of the χ^2-distribution

k	$\chi^2_{0.005}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
1	3.93×10^{-5}	9.82E−4	3.93E−3	2.706	3.841	5.024	6.635	7.879
2	0.0100	0.0506	0.103	4.605	5.991	7.378	9.210	10.597
3	0.0717	0.216	0.352	6.251	7.815	9.348	11.345	12.838
4	0.207	0.484	0.711	7.779	9.488	11.143	13.277	14.860
5	0.412	0.831	1.145	9.236	11.070	12.832	15.086	16.750
6	0.676	1.237	1.635	10.645	12.592	14.449	16.812	18.548
7	0.989	1.690	2.167	12.017	14.067	16.013	18.475	20.278
8	1.344	2.180	2.733	13.362	15.507	17.535	20.090	21.955
9	1.735	2.700	3.325	14.684	16.919	19.023	21.666	23.589
10	2.156	3.247	3.940	15.987	18.307	20.483	23.209	25.188
11	2.603	3.816	4.575	17.275	19.675	21.920	24.725	26.757
12	3.074	4.404	5.226	18.549	21.026	23.336	26.217	28.300
13	3.565	5.009	5.892	19.812	22.362	24.736	27.688	29.819
14	4.075	5.629	6.571	21.064	23.685	26.119	29.141	31.319
15	4.601	6.262	7.261	22.307	24.996	27.488	30.578	32.801
16	5.142	6.908	7.962	23.542	26.296	28.845	32.000	34.267
17	5.697	7.564	8.672	24.769	27.587	30.191	33.409	35.718
18	6.265	8.231	9.390	25.989	28.869	31.526	34.805	37.156
19	6.844	8.907	10.117	27.204	30.144	32.852	36.191	38.582
20	7.434	9.591	10.851	28.412	31.410	34.170	37.566	39.997
21	8.034	10.283	11.591	29.615	32.671	35.479	38.932	41.401
22	8.643	10.982	12.338	30.813	33.924	36.781	40.289	42.796
23	9.260	11.688	13.091	32.007	35.172	38.076	41.638	44.181
24	9.886	12.401	13.848	33.196	36.415	39.364	42.980	45.558
25	10.520	13.120	14.611	34.382	37.652	40.646	44.314	46.928
26	11.160	13.844	15.379	35.563	38.885	41.923	45.642	48.290
27	11.808	14.573	16.151	36.741	40.113	43.194	46.963	49.645
28	12.461	15.308	16.928	37.916	41.337	44.461	48.278	50.993
29	13.121	16.047	17.708	39.087	42.557	45.722	49.588	52.336
30	13.787	16.791	18.493	40.256	43.773	46.979	50.892	53.672
35	17.192	20.569	22.465	46.059	49.802	53.203	57.342	60.275
40	20.707	24.433	26.509	51.805	55.758	59.342	63.691	66.766
45	24.311	28.366	30.612	57.505	61.656	65.410	69.957	73.166
50	27.991	32.357	34.764	63.167	67.505	71.420	76.154	79.490
60	35.535	40.482	43.188	74.397	79.082	83.298	88.379	91.952
70	43.275	48.758	51.739	85.527	90.531	95.023	100.425	104.215
80	51.172	57.153	60.391	96.578	101.879	106.629	112.329	116.321
90	59.196	65.647	69.126	107.565	113.145	118.136	124.116	128.299
100	67.328	74.222	77.929	118.498	124.342	129.561	135.807	140.169

Table 4A. F-values for 0.5% significance level with numerator degrees of freedom N and denominator degrees of freedom D

	N								
D	1	2	3	4	5	6	7	8	9
1	16211	20000	21615	25000	23056	23437	23715	23925	24091
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62

Table 4B. F-values for 1% significance level with numerator degrees of freedom N and denominator degrees of freedom D

	N								
D	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Table 4C. F-values for 2.5% significance level with numerator degrees of freedom N and denominator degrees of freedom D

	N								
D	1	2	3	4	5	6	7	8	9
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

Table 4D. F-values for 5% significance level with numerator degrees of freedom N and denominator degrees of freedom D

	N								
D	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Table 4E. F-values for 10% significance level with numerator degrees of freedom N and denominator degrees of freedom D

	N								
D	1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

Table 5A

$n = 5$	
T	P
0	0.0313
1	0.0625
2	0.0938
3	0.1563
4	0.2188
5	0.3125
6	0.4063
7	0.5000

$n = 6$	
T	P
0	0.0156
1	0.0313
2	0.0469
3	0.0781
4	0.1094
5	0.1563
6	0.2188
7	0.2813
8	0.3438
9	0.4219
10	0.5000

$n = 7$	
T	P
0	0.0078
1	0.0156
2	0.0234
3	0.0391
4	0.0547
5	0.0781
6	0.1094
7	0.1484
8	0.1875
9	0.2344
10	0.2891
11	0.3438
12	0.4063
13	0.4688
14	0.5313

$n = 8$	
T	P
0	0.0039
1	0.0078
2	0.0117
3	0.0195
4	0.0273
5	0.0391
6	0.0547
7	0.0742
8	0.0977
9	0.1250
10	0.1563
11	0.1914
12	0.2305
13	0.2734
14	0.3203
15	0.3711
16	0.4219
17	0.4727
18	0.5273

$n = 9$	
T	P
0	0.0020
1	0.0039
2	0.0059
3	0.0098
4	0.0137
5	0.0195
6	0.0273
7	0.0371
8	0.0488
9	0.0645
10	0.0820
11	0.1016
12	0.1250
13	0.1504
14	0.1797
15	0.2129
16	0.2480
17	0.2852
18	0.3262
19	0.3672
20	0.4102
21	0.4551
22	0.5000

$n = 10$	
T	P
0	0.0010
1	0.0020
2	0.0029
3	0.0049
4	0.0068
5	0.0098
6	0.0137
7	0.0186
8	0.0244
9	0.0322
10	0.0420
11	0.0527
12	0.0654
13	0.0801
14	0.0967
15	0.1162
16	0.1377
17	0.1611
18	0.1875
19	0.2158
20	0.2461
21	0.2783
22	0.3125
23	0.3477
24	0.3848
25	0.4229
26	0.4609
27	0.5000

$n = 11$	
T	P
0	0.0005
1	0.0010
2	0.0015
3	0.0024
4	0.0034
5	0.0049
6	0.0068
7	0.0093
8	0.0122
9	0.0161
10	0.0210
11	0.0269
12	0.0337
13	0.0415
14	0.0508
15	0.0615
16	0.0737
17	0.0874
18	0.1030
19	0.1201
20	0.1392
21	0.1602
22	0.1826
23	0.2065
24	0.2324
25	0.2598
26	0.2886
27	0.3188
28	0.3501
29	0.3823
30	0.4155
31	0.4492
32	0.4829
33	0.5171

$n = 12$	
T	P
0	0.0002
1	0.0005
2	0.0007
3	0.0012
4	0.0017
5	0.0024
6	0.0034
7	0.0046
8	0.0061
9	0.0081
10	0.0105
11	0.0134
12	0.0171
13	0.0212
14	0.0261
15	0.0320
16	0.0386
17	0.0461
18	0.0549
19	0.0647
20	0.0757
21	0.0881
22	0.1018
23	0.1167
24	0.1331
25	0.1506
26	0.1697
27	0.1902
28	0.2119
29	0.2349
30	0.2593
31	0.2847
32	0.3110
33	0.3386
34	0.3667
35	0.3955

Table 5B

$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 15$	
T	P	T	P	T	P	T	P	T	P
36	0.4250	33	0.2072	25	0.0453	12	0.0021	50	0.2997
37	0.4548	34	0.2274	26	0.0520	13	0.0027	51	0.3193
38	0.4849	35	0.2487	27	0.0594	14	0.0034	52	0.3394
39	0.5151	36	0.2709	28	0.0676	15	0.0042	53	0.3599
$n = 13$		37	0.2939	29	0.0765	16	0.0051	54	0.3808
0	0.0001	38	0.3177	30	0.0863	17	0.0062	55	0.4020
1	0.0002	39	0.3424	31	0.0969	18	0.0075	56	0.4235
2	0.0004	40	0.3677	32	0.1083	19	0.0090	57	0.4452
3	0.0006	41	0.3934	33	0.1206	20	0.0108	58	0.4670
4	0.0009	42	0.4197	34	0.1338	21	0.0128	59	0.4890
5	0.0012	43	0.4463	35	0.1479	22	0.0151	60	0.5110
6	0.0017	44	0.4730	36	0.1629	23	0.0177	$n = 16$	
7	0.0023	45	0.5000	37	0.1788	24	0.0206	3	0.0001
8	0.0031	$n = 14$		38	0.1955	25	0.0240	5	0.0002
9	0.0040	0	0.0001	39	0.2131	26	0.0277	7	0.0003
10	0.0052	2	0.0002	40	0.2316	27	0.0319	8	0.0004
11	0.0067	3	0.0003	41	0.2508	28	0.0365	9	0.0005
12	0.0085	4	0.0004	42	0.2708	29	0.0416	10	0.0007
13	0.0107	5	0.0006	43	0.2915	30	0.0473	11	0.0008
14	0.0133	6	0.0009	44	0.3129	31	0.0535	12	0.0011
15	0.0164	7	0.0012	45	0.3349	32	0.0603	13	0.0013
16	0.0199	8	0.0015	46	0.3574	33	0.0677	14	0.0017
17	0.0239	9	0.0020	47	0.3804	34	0.0757	15	0.0021
18	0.0287	10	0.0026	48	0.4039	35	0.0844	16	0.0026
19	0.0341	11	0.0034	49	0.4276	36	0.0938	17	0.0031
20	0.0402	12	0.0043	50	0.4516	37	0.1039	18	0.0038
21	0.0471	13	0.0054	51	0.4758	38	0.1147	19	0.0046
22	0.0549	14	0.0067	52	0.5000	39	0.1262	20	0.0055
23	0.0636	15	0.0083	$n = 15$		40	0.1384	21	0.0065
24	0.0732	16	0.0101	0	0.0001	41	0.1514	22	0.0078
25	0.0839	17	0.0123	3	0.0002	42	0.1651	23	0.0091
26	0.0955	18	0.0148	5	0.0003	43	0.1796	24	0.0107
27	0.1082	19	0.0176	6	0.0004	44	0.1947	25	0.0125
28	0.1219	20	0.0209	7	0.0006	45	0.2106	26	0.0145
29	0.1367	21	0.0247	8	0.0008	46	0.2271	27	0.0168
30	0.1527	22	0.0290	9	0.0010	47	0.2444	28	0.0193
31	0.1698	23	0.0338	10	0.0013	48	0.2622	29	0.0222
32	0.1879	24	0.0392	11	0.0017	49	0.2807	30	0.0253

Table 5C

$n = 16$		$n = 17$		$n = 17$		$n = 18$		$n = 18$	
T	P	T	P	T	P	T	P	T	P
31	0.0288	4	0.0001	46	0.0797	17	0.0008	55	0.0982
32	0.0327	8	0.0002	47	0.0871	18	0.0010	56	0.1061
33	0.0370	9	0.0003	48	0.0950	19	0.0012	57	0.1144
34	0.0416	11	0.0004	49	0.1034	20	0.0014	58	0.1231
35	0.0467	12	0.0005	50	0.1123	21	0.0017	59	0.1323
36	0.0523	13	0.0007	51	0.1218	22	0.0020	60	0.1419
37	0.0583	14	0.0008	52	0.1317	23	0.0024	61	0.1519
38	0.0649	15	0.0010	53	0.1421	24	0.0028	62	0.1624
39	0.0719	16	0.0013	54	0.1530	25	0.0033	63	0.1733
40	0.0795	17	0.0016	55	0.1645	26	0.0038	64	0.1846
41	0.0877	18	0.0019	56	0.1764	27	0.0045	65	0.1964
42	0.0964	19	0.0023	57	0.1889	28	0.0052	66	0.2086
43	0.1057	20	0.0028	58	0.2019	29	0.0060	67	0.2211
44	0.1156	21	0.0033	59	0.2153	30	0.0069	68	0.2341
45	0.1261	22	0.0040	60	0.2293	31	0.0080	69	0.2475
46	0.1372	23	0.0047	61	0.2437	32	0.0091	70	0.2613
47	0.1489	24	0.0055	62	0.2585	33	0.0104	71	0.2754
48	0.1613	25	0.0064	63	0.2738	34	0.0118	72	0.2899
49	0.1742	26	0.0075	64	0.2895	35	0.0134	73	0.3047
50	0.1877	27	0.0087	65	0.3056	36	0.0152	74	0.3198
51	0.2019	28	0.0101	66	0.3221	37	0.0171	75	0.3353
52	0.2166	29	0.0116	67	0.3389	38	0.0192	76	0.3509
53	0.2319	30	0.0133	68	0.3559	39	0.0216	77	0.3669
54	0.2477	31	0.0153	69	0.3733	40	0.0241	78	0.3830
55	0.2641	32	0.0174	70	0.3910	41	0.0269	79	0.3994
56	0.2809	33	0.0198	71	0.4088	42	0.0300	80	0.4159
57	0.2983	34	0.0224	72	0.4268	43	0.0333	81	0.4325
58	0.3161	35	0.0253	73	0.4450	44	0.0368	82	0.4493
59	0.3343	36	0.0284	74	0.4633	45	0.0407	83	0.4661
60	0.3529	37	0.0319	75	0.4816	46	0.0449	84	0.4831
61	0.3718	38	0.0357	76	0.5000	47	0.0494	85	0.5000
62	0.3910	39	0.0398	$n = 18$		48	0.0542	$n = 19$	
63	0.4104	40	0.0443	6	0.0001	49	0.0594	9	0.0001
64	0.4301	41	0.0492	10	0.0002	50	0.0649	13	0.0002
65	0.4500	42	0.0544	12	0.0003	51	0.0708	15	0.0003
66	0.4699	43	0.0601	14	0.0004	52	0.0770	17	0.0004
67	0.4900	44	0.0662	15	0.0005	53	0.0837	18	0.0005
68	0.5100	45	0.0727	16	0.0006	54	0.0907	19	0.0006

Table 5D

$n = 19$		$n = 19$		$n = 20$		$n = 20$		$n = 20$	
T	P	T	P	T	P	T	P	T	P
20	0.0007	58	0.0723	11	0.0001	56	0.0348	94	0.3506
21	0.0008	59	0.0782	16	0.0002	57	0.0379	95	0.3643
22	0.0010	60	0.0844	19	0.0003	58	0.0413	96	0.3781
23	0.0012	61	0.0909	20	0.0004	59	0.0448	97	0.3921
24	0.0014	62	0.0978	22	0.0005	60	0.0487	98	0.4062
25	0.0017	63	0.1051	23	0.0006	61	0.0527	99	0.4204
26	0.0020	64	0.1127	24	0.0007	62	0.0570	100	0.4347
27	0.0023	65	0.1206	25	0.0008	63	0.0615	101	0.4492
28	0.0027	66	0.1290	26	0.0010	64	0.0664	102	0.4636
29	0.0031	67	0.1377	27	0.0012	65	0.0715	103	0.4782
30	0.0036	68	0.1467	28	0.0014	66	0.0768	104	0.4927
31	0.0041	69	0.1562	29	0.0016	67	0.0825	105	0.5073
32	0.0047	70	0.1660	30	0.0018	68	0.0884	$n = 21$	
33	0.0054	71	0.1762	31	0.0021	69	0.0947	14	0.0001
34	0.0062	72	0.1868	32	0.0024	70	0.1012	20	0.0002
35	0.0070	73	0.1977	33	0.0028	71	0.1081	22	0.0003
36	0.0080	74	0.2090	34	0.0032	72	0.1153	24	0.0004
37	0.0090	75	0.2207	35	0.0036	73	0.1227	26	0.0005
38	0.0102	76	0.2327	36	0.0042	74	0.1305	27	0.0006
39	0.0115	77	0.2450	37	0.0047	75	0.1387	28	0.0007
40	0.0129	78	0.2576	38	0.0053	76	0.1471	29	0.0008
41	0.0145	79	0.2706	39	0.0060	77	0.1559	30	0.0009
42	0.0162	80	0.2839	40	0.0068	78	0.1650	31	0.0011
43	0.0180	81	0.2974	41	0.0077	79	0.1744	32	0.0012
44	0.0201	82	0.3113	42	0.0086	80	0.1841	33	0.0014
45	0.0223	83	0.3254	43	0.0096	81	0.1942	34	0.0016
46	0.0247	84	0.3397	44	0.0107	82	0.2045	35	0.0019
47	0.0273	85	0.3543	45	0.0120	83	0.2152	36	0.0021
48	0.0301	86	0.3690	46	0.0133	84	0.2262	37	0.0024
49	0.0331	87	0.3840	47	0.0148	85	0.2375	38	0.0028
50	0.0364	88	0.3991	48	0.0164	86	0.2490	39	0.0031
51	0.0399	89	0.4144	49	0.0181	87	0.2608	40	0.0036
52	0.0437	90	0.4298	50	0.0200	88	0.2729	41	0.0040
53	0.0478	91	0.4453	51	0.0220	89	0.2853	42	0.0045
54	0.0521	92	0.4609	52	0.0242	90	0.2979	43	0.0051
55	0.0567	93	0.4765	53	0.0266	91	0.3108	44	0.0057
56	0.0616	94	0.4922	54	0.0291	92	0.3238	45	0.0063
57	0.0668	95	0.5078	55	0.0319	93	0.3371	46	0.0071

Table 5E

$n = 21$		$n = 21$		$n = 22$		$n = 22$		$n = 22$	
T	P	T	P	T	P	T	P	T	P
47	0.0079	85	0.1519	33	0.0007	71	0.0369	109	0.2940
48	0.0088	86	0.1602	34	0.0008	72	0.0397	110	0.3051
49	0.0097	87	0.1688	35	0.0010	73	0.0427	111	0.3164
50	0.0108	88	0.1777	36	0.0011	74	0.0459	112	0.3278
51	0.0119	89	0.1869	37	0.0013	75	0.0492	113	0.3394
52	0.0132	90	0.1963	38	0.0014	76	0.0527	114	0.3512
53	0.0145	91	0.2060	39	0.0016	77	0.0564	115	0.3631
54	0.0160	92	0.2160	40	0.0018	78	0.0603	116	0.3751
55	0.0175	93	0.2262	41	0.0021	79	0.0644	117	0.3873
56	0.0192	94	0.2367	42	0.0023	80	0.0687	118	0.3995
57	0.0210	95	0.2474	43	0.0026	81	0.0733	119	0.4119
58	0.0230	96	0.2584	44	0.0030	82	0.0780	120	0.4243
59	0.0251	97	0.2696	45	0.0033	83	0.0829	121	0.4368
60	0.0273	98	0.2810	46	0.0037	84	0.0881	122	0.4494
61	0.0298	99	0.2927	47	0.0042	85	0.0935	123	0.4620
62	0.0323	100	0.3046	48	0.0046	86	0.0991	124	0.4746
63	0.0351	101	0.3166	49	0.0052	87	0.1050	125	0.4873
64	0.0380	102	0.3289	50	0.0057	88	0.1111	126	0.5000
65	0.0411	103	0.3414	51	0.0064	89	0.1174	$n = 23$	
66	0.0444	104	0.3540	52	0.0070	90	0.1240	21	0.0001
67	0.0479	105	0.3667	53	0.0078	91	0.1308	28	0.0002
68	0.0516	106	0.3796	54	0.0086	92	0.1378	31	0.0003
69	0.0555	107	0.3927	55	0.0095	93	0.1451	33	0.0004
70	0.0597	108	0.4058	56	0.0104	94	0.1527	35	0.0005
71	0.0640	109	0.4191	57	0.0115	95	0.1604	36	0.0006
72	0.0686	110	0.4324	58	0.0126	96	0.1685	38	0.0007
73	0.0735	111	0.4459	59	0.0138	97	0.1767	39	0.0008
74	0.0786	112	0.4593	60	0.0161	98	0.1853	40	0.0009
75	0.0839	113	0.4729	61	0.0164	99	0.1940	41	0.0011
76	0.0895	114	0.4864	62	0.0179	100	0.2030	42	0.0012
77	0.0953	115	0.5000	63	0.0195	101	0.2122	43	0.0014
78	0.1015	$n = 22$		64	0.0212	102	0.2217	44	0.0015
79	0.1078	18	0.0001	65	0.0231	103	0.2314	45	0.0017
80	0.1145	23	0.0002	66	0.0250	104	0.2413	46	0.0019
81	0.1214	26	0.0003	67	0.0271	105	0.2514	47	0.0022
82	0.1286	29	0.0004	68	0.0293	106	0.2618	48	0.0024
83	0.1361	30	0.0005	69	0.0317	107	0.2723	49	0.0027
84	0.1439	32	0.0006	70	0.0342	108	0.2830	50	0.0030

Table 5F

$n = 23$		$n = 23$		$n = 23$		$n = 24$		$n = 24$	
T	P	T	P	T	P	T	P	T	P
51	0.0034	89	0.0712	127	0.3770	62	0.0053	100	0.0800
52	0.0037	90	0.0755	128	0.3884	63	0.0058	101	0.0844
53	0.0041	91	0.0801	129	0.3999	64	0.0063	102	0.0890
54	0.0046	92	0.0848	130	0.4115	65	0.0069	103	0.0938
55	0.0051	93	0.0897	131	0.4231	66	0.0075	104	0.0987
56	0.0056	94	0.0948	132	0.4348	67	0.0082	105	0.1038
57	0.0061	95	0.1001	133	0.4466	68	0.0089	106	0.1091
58	0.0068	96	0.1056	134	0.4584	69	0.0097	107	0.1146
59	0.0074	97	0.1113	135	0.4703	70	0.0106	108	0.1203
60	0.0082	98	0.1172	136	0.4822	71	0.0115	109	0.1261
61	0.0089	99	0.1234	137	0.4941	72	0.0124	110	0.1322
62	0.0098	100	0.1297	138	0.5060	73	0.0135	111	0.1384
63	0.0107	101	0.1363	$n = 24$		74	0.0146	112	0.1448
64	0.0117	102	0.1431	25	0.0001	75	0.0157	113	0.1515
65	0.0127	103	0.1501	32	0.0002	76	0.0170	114	0.1583
66	0.0138	104	0.1573	36	0.0003	77	0.0183	115	0.1653
67	0.0150	105	0.1647	38	0.0004	78	0.0197	116	0.1724
68	0.0163	106	0.1723	40	0.0005	79	0.0212	117	0.1798
69	0.0177	107	0.1802	42	0.0006	80	0.0228	118	0.1874
70	0.0192	108	0.1883	43	0.0007	81	0.0245	119	0.1951
71	0.0208	109	0.1965	44	0.0008	82	0.0263	120	0.2031
72	0.0224	110	0.2050	45	0.0009	83	0.0282	121	0.2112
73	0.0242	111	0.2137	46	0.0010	84	0.0302	122	0.2195
74	0.0261	112	0.2226	47	0.0011	85	0.0323	123	0.2279
75	0.0281	113	0.2317	48	0.0013	86	0.0346	124	0.2366
76	0.0303	114	0.2410	49	0.0014	87	0.0369	125	0.2454
77	0.0325	115	0.2505	50	0.0016	88	0.0394	126	0.2544
78	0.0349	116	0.2601	51	0.0018	89	0.0420	127	0.2635
79	0.0374	117	0.2700	52	0.0020	90	0.0447	128	0.2728
80	0.0401	118	0.2800	53	0.0022	91	0.0475	129	0.2823
81	0.0429	119	0.2902	54	0.0024	92	0.0505	130	0.2919
82	0.0459	120	0.3005	55	0.0027	93	0.0537	131	0.3017
83	0.0490	121	0.3110	56	0.0029	94	0.0570	132	0.3115
84	0.0523	122	0.3217	57	0.0033	95	0.0604	133	0.3216
85	0.0557	123	0.3325	58	0.0036	96	0.0640	134	0.3317
86	0.0593	124	0.3434	59	0.0040	97	0.0678	135	0.3420
87	0.0631	125	0.3545	60	0.0044	98	0.0717	136	0.3524
88	0.0671	126	0.3657	61	0.0048	99	0.0758	137	0.3629

Table 5G

$n = 24$		$n = 25$		$n = 25$		$n = 25$	
T	P	T	P	T	P	T	P
138	0.3735	66	0.0040	100	0.0479	136	0.2474
139	0.3841	67	0.0044	101	0.0507	137	0.2539
140	0.3949	68	0.0048	102	0.0537	138	0.2625
141	0.4058	69	0.0053	103	0.0567	139	0.2712
142	0.4167	70	0.0057	104	0.0600	140	0.2801
143	0.4277	71	0.0062	105	0.0633	141	0.2891
144	0.4387	72	0.0068	106	0.0668	142	0.2983
145	0.4498	70	0.0057	107	0.0705	143	0.3075
146	0.4609	71	0.0062	108	0.0742	144	0.3169
147	0.4721	72	0.0068	109	0.0782	145	0.3264
148	0.4832	73	0.0074	110	0.0822	146	0.3360
149	0.4944	74	0.0080	111	0.0865	147	0.3458
150	0.5056	75	0.0087	112	0.0909	148	0.3556
$n = 25$		76	0.0094	113	0.0954	149	0.3655
29	0.0001	77	0.0101	114	0.1001	150	0.3755
37	0.0002	78	0.0110	115	0.1050	151	0.3856
41	0.0003	79	0.0118	116	0.1100	152	0.3957
43	0.0004	80	0.0128	117	0.1152	153	0.4060
45	0.0005	81	0.0137	118	0.1205	154	0.4163
47	0.0006	82	0.0148	119	0.1261	155	0.4266
48	0.0007	83	0.0159	120	0.1317	156	0.4370
50	0.0008	84	0.0171	121	0.1376	157	0.4474
51	0.0009	85	0.0183	122	0.1436	158	0.4579
52	0.0010	86	0.0197	123	0.1498	159	0.4684
53	0.0011	87	0.0211	124	0.1562	160	0.4789
54	0.0013	88	0.0226	125	0.1627	161	0.4895
55	0.0014	89	0.0241	126	0.1694	162	0.5000
56	0.0015	90	0.0258	127	0.1763		
57	0.0017	91	0.0275	128	0.1833		
58	0.0019	92	0.0294	129	0.1905		
59	0.0021	93	0.0313	130	0.1979		
60	0.0023	94	0.0334	131	0.2054		
61	0.0025	95	0.0355	132	0.2131		
62	0.0028	96	0.0377	133	0.2209		
63	0.0031	97	0.0401	134	0.2289		
64	0.0034	98	0.0426	135	0.2371		
65	0.0037	99	0.0452	136	0.2454		

Table 6. Quantities of Mann-Whitney U test statistic

		m													
n	p	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	.005	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	.01	0	0	0	0	0	0	0	0	0	0	0	1	1	1
	.025	0	0	0	0	0	0	1	1	1	1	2	2	2	2
	.05	0	0	0	1	1	1	2	2	2	2	3	3	4	4
	.10	0	1	1	2	2	2	3	3	4	4	5	5	5	6
	.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	.005	0	0	0	0	0	0	0	1	1	1	2	2	2	3
3	.01	0	0	0	0	0	1	1	2	2	2	3	3	3	4
	.025	0	0	0	1	2	2	3	3	4	4	5	5	6	6
	.05	0	1	1	2	3	3	4	5	5	6	6	7	8	8
	.10	1	2	2	3	4	5	6	6	7	8	9	10	11	11
	.001	0	0	0	0	0	0	0	0	1	1	1	2	2	2
	.005	0	0	0	0	1	1	2	2	3	3	4	4	5	6
4	.01	0	0	0	1	2	2	3	4	4	5	6	6	7	9
	.025	0	0	1	2	3	4	5	5	6	7	8	9	10	11
	.05	0	1	2	3	4	5	6	7	8	9	10	11	12	13
	.10	1	2	4	5	6	7	8	10	11	12	13	14	16	17
	.001	0	0	0	0	0	0	1	2	2	3	3	4	4	5
	.005	0	0	0	1	2	2	3	4	5	6	7	8	8	9
5	.01	0	0	1	2	3	4	5	6	7	8	9	10	11	12
	.025	0	1	2	3	4	6	7	8	9	10	12	13	14	15
	.05	1	2	3	5	6	7	9	10	12	13	14	16	17	19
	.10	2	3	5	6	8	9	11	13	14	16	18	19	21	23
	.001	0	0	0	0	0	0	2	3	4	5	5	6	7	8
	.005	0	0	1	2	3	4	5	6	7	8	10	11	12	13
6	.01	0	0	2	3	4	5	7	8	9	10	12	13	14	16
	.025	0	2	3	4	6	7	9	11	12	14	15	17	18	20
	.05	1	3	4	6	8	9	11	13	15	17	18	20	22	24
	.10	2	4	6	8	10	12	14	16	18	20	22	24	26	28
	.001	0	0	0	0	1	2	3	4	6	7	8	9	10	11
	.005	0	0	1	2	4	5	7	8	10	11	13	14	16	17
7	.01	0	1	2	4	5	7	8	10	12	13	15	17	18	20
	.025	0	2	4	6	7	9	11	13	15	17	19	21	23	25
	.05	1	3	5	7	9	12	14	16	18	20	22	25	27	29
	.10	2	5	7	9	12	14	17	19	22	24	27	29	32	34

Table 6 continued

		m													
n	*p*	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	.001	0	0	0	1	2	3	5	6	7	9	10	12	13	15
	.005	0	0	2	3	5	7	8	10	12	14	16	18	19	21
8	.01	0	1	3	5	7	8	10	12	14	16	18	21	23	25
	.025	1	3	5	7	9	11	14	16	18	20	23	25	27	30
	.05	2	4	6	9	11	14	16	19	21	24	27	29	32	34
	.10	3	6	8	11	14	17	20	23	25	28	31	34	37	40
	.001	0	0	0	2	3	4	6	8	9	11	13	15	16	18
	.005	0	1	2	4	6	8	10	12	14	17	19	21	23	25
9	.01	0	2	4	6	8	10	12	15	17	19	22	24	27	29
	.025	1	3	5	8	11	13	16	18	21	24	27	29	32	35
	.05	2	5	7	10	13	16	19	22	25	28	31	34	37	40
	.10	3	6	10	13	16	19	23	26	29	32	36	39	42	46
	.001	0	0	1	2	4	6	7	9	11	13	15	18	20	22
	.005	0	1	3	5	7	10	12	14	17	19	22	25	27	30
10	.01	0	2	4	7	9	12	14	17	20	23	25	28	31	34
	.025	1	4	6	9	12	15	18	21	24	27	30	34	37	40
	.05	2	5	8	12	15	18	21	25	28	32	35	38	42	45
	.10	4	7	11	14	18	22	25	29	33	37	40	44	48	52
	.001	0	0	1	3	5	7	9	11	13	16	18	21	23	25
	.005	0	1	3	6	8	11	14	17	19	22	25	28	31	34
11	.01	0	2	5	8	10	13	16	19	23	26	29	32	35	38
	.025	1	4	7	10	14	17	20	24	27	31	34	38	41	45
	.05	2	6	9	13	17	20	24	28	32	35	39	43	47	51
	.10	4	8	12	16	20	24	28	32	37	41	45	49	53	58
	.001	0	0	1	3	5	8	10	13	15	18	21	24	26	29
	.005	0	2	4	7	10	13	16	19	22	25	28	32	35	38
12	.01	0	3	6	9	12	15	18	22	25	29	32	36	39	43
	.025	2	5	8	12	15	19	23	27	30	34	38	42	46	50
	.05	3	6	10	14	18	22	27	31	35	39	43	48	52	56
	.10	5	9	13	18	22	27	31	36	40	45	50	54	59	64
	.001	0	0	2	4	6	9	12	15	18	21	24	27	30	33
	.005	0	2	4	8	11	14	18	21	25	28	32	35	39	43
13	.01	1	3	6	10	13	17	21	24	28	32	36	40	44	48
	.025	2	5	9	13	17	21	25	29	34	38	42	46	51	55
	.05	3	7	11	16	20	25	29	34	38	43	48	52	57	62
	.10	5	10	14	19	24	29	34	39	44	49	54	59	64	69

Table 6 continued

		m													
n	p	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	.001	0	0	2	4	7	10	13	16	20	23	26	30	33	37
	.005	0	2	5	8	12	16	19	23	27	31	35	39	43	47
14	.01	1	3	7	11	14	18	23	27	31	35	39	44	48	52
	.025	2	6	10	14	18	23	27	32	37	41	46	51	56	60
	.05	4	8	12	17	22	27	32	37	42	47	52	57	62	67
	.10	5	11	16	21	26	32	37	42	48	53	59	64	70	75
	.001	0	0	2	5	8	11	15	18	22	25	29	33	37	41
	.005	0	3	6	9	13	17	21	25	30	34	38	43	47	52
15	.01	1	4	8	12	16	20	25	29	34	38	43	48	52	57
	.025	2	6	11	15	20	25	30	35	40	45	50	55	60	65
	.05	4	8	13	19	24	29	34	40	45	51	56	62	67	73
	.10	6	11	17	23	28	34	40	46	52	58	64	69	75	81
	.001	0	0	3	6	9	12	16	20	24	28	32	36	40	44
	.005	0	3	6	10	14	19	23	28	32	37	42	46	51	56
16	.01	1	4	8	13	17	22	27	32	37	42	47	52	57	62
	.025	2	7	12	16	22	27	32	38	43	48	54	60	65	71
	.05	4	9	15	20	26	31	37	43	49	55	61	66	72	78
	.10	6	12	18	24	30	37	43	49	55	62	68	75	81	87
	.001	0	1	3	6	10	14	18	22	26	30	35	39	44	48
	.005	0	3	7	11	16	20	25	30	35	40	45	50	55	61
17	.01	1	5	9	14	19	24	29	34	39	45	50	56	61	67
	.025	3	7	12	18	23	29	35	40	46	52	58	64	70	76
	.05	4	10	16	21	27	34	40	46	52	58	65	71	78	84
	.10	7	13	19	26	32	39	46	53	59	66	73	80	86	93
	.001	0	1	4	7	11	15	19	24	28	33	38	43	47	52
	.005	0	3	7	12	17	22	27	32	38	43	48	54	59	65
18	.01	1	5	10	15	20	25	31	37	42	48	54	60	66	71
	.025	3	8	13	19	25	31	37	43	49	56	62	68	75	81
	.05	5	10	17	23	29	36	42	49	56	62	69	76	83	89
	.10	7	14	21	28	35	42	49	56	63	70	78	85	92	99
	.001	0	1	4	8	12	16	21	26	30	35	41	46	51	56
	.005	1	4	8	13	18	23	29	34	40	46	52	58	64	70
19	.01	2	5	10	16	21	27	33	39	45	51	57	64	70	76
	.025	3	8	14	20	26	33	39	46	53	59	66	73	79	86
	.05	5	11	18	24	31	38	45	52	59	66	73	81	88	95
	.10	8	15	22	29	37	44	52	59	67	74	82	90	98	105

Table 7 A. Cumulative binomial probability distribution for $n = 5$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.951	.904	.859	.815	.774	.734	.696	.659	.624	.590
1	.999	.996	.992	.985	.977	.968	.958	.946	.933	.919
2	1.00	1.00	1.00	.999	.999	.998	.997	.995	.994	.991
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.558	.528	.498	.470	.444	.418	.394	.371	.349	.328
1	.903	.888	.871	.853	.835	.817	.797	.778	.758	.737
2	.989	.986	.982	.978	.973	.968	.963	.956	.949	.942
3	.999	.999	.999	.998	.998	.997	.996	.996	.994	.993
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.308	.289	.271	.254	.237	.222	.207	.193	.180	.168
1	.717	.696	.675	.654	.633	.612	.591	.570	.549	.528
2	.934	.926	.916	.907	.896	.886	.874	.862	.850	.837
3	.992	.990	.989	.987	.984	.982	.979	.976	.973	.969
4	1.00	.999	.999	.999	.999	.999	.999	.998	.998	.998

	p									
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.156	.145	.135	.125	.116	.107	.099	.092	.084	.078
1	.508	.487	.468	.448	.428	.409	.391	.372	.354	.337
2	.823	.809	.795	.780	.765	.749	.733	.717	.700	.683
3	.965	.961	.956	.951	.946	.940	.934	.927	.920	.913
4	.997	.997	.996	.995	.995	.994	.993	.992	.991	.990

	p									
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.071	.066	.060	.055	.050	.046	.042	.038	.035	.031
1	.320	.303	.287	.271	.256	.241	.227	.213	.200	.187
2	.665	.647	.630	.611	.593	.575	.556	.537	.519	.500
3	.905	.897	.888	.879	.869	.859	.848	.837	.825	.812
4	.988	.987	.985	.984	.982	.979	.977	.975	.972	.969

Table 7 B. Cumulative binomial probability distribution for $n = 6$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.941	.886	.833	.783	.735	.690	.647	.606	.568	.531
1	.999	.994	.988	.978	.967	.954	.939	.923	.905	.886
2	1.00	1.00	.999	.999	.998	.996	.994	.991	.988	.984
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.497	.464	.434	.405	.377	.351	.327	.304	.282	.262
1	.866	.844	.822	.800	.776	.753	.729	.704	.680	.655
2	.979	.974	.968	.961	.953	.944	.934	.924	.913	.901
3	.998	.997	.997	.995	.994	.993	.991	.988	.986	.983
4	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.243	.225	.208	.193	.178	.164	.151	.139	.128	.118
1	.631	.606	.582	.558	.534	.510	.487	.464	.442	.420
2	.888	.875	.861	.846	.831	.814	.798	.780	.763	.744
3	.980	.976	.972	.967	.962	.957	.951	.944	.937	.930
4	.998	.997	.997	.996	.995	.994	.993	.992	.991	.989
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999

	p									
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.108	.099	.090	.083	.075	.069	.063	.057	.052	.047
1	.399	.378	.358	.338	.319	.301	.283	.266	.249	.233
2	.726	.706	.687	.667	.647	.627	.606	.586	.565	.544
3	.921	.913	.903	.893	.883	.871	.860	.847	.834	.821
4	.987	.985	.983	.980	.978	.975	.971	.968	.963	.959
5	.999	.999	.999	.998	.998	.998	.997	.997	.996	.996

	p									
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.042	.038	.034	.031	.028	.025	.022	.020	.018	.016
1	.218	.203	.190	.176	.164	.152	.140	.129	.119	.109
2	.524	.503	.482	.462	.442	.421	.402	.382	.363	.344
3	.807	.792	.777	.761	.745	.728	.711	.693	.675	.656
4	.954	.949	.943	.937	.931	.924	.916	.908	.900	.891
5	.995	.995	.994	.993	.992	.991	.989	.988	.986	.984

Table 7 C. Cumulative binomial probability distribution for $n = 7$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.932	0.868	.808	.751	.698	.648	.602	.558	.517	.478
1	.998	.992	.983	.971	.956	.938	.919	.897	.875	.850
2	1.00	1.00	.999	.998	.996	.994	.990	.986	.981	.974
3	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.997
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.442	.409	.377	.348	.321	.295	.271	.249	.229	.210
1	.825	.799	.772	.744	.717	.689	.660	.632	.604	.577
2	.967	.958	.949	.938	.926	.913	.899	.885	.869	.852
3	.996	.995	.993	.991	.988	.985	.981	.977	.972	.967
4	1.00	1.00	.999	.999	.999	.998	.998	.997	.996	.995
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.192	.176	.160	.146	.133	.122	.110	.100	.091	.082
1	.549	.522	.496	.470	.445	.420	.396	.373	.351	.329
2	.834	.816	.797	.777	.756	.735	.714	.692	.670	.647
3	.961	.954	.946	.938	.929	.920	.909	.898	.887	.874
4	.994	.993	.991	.989	.987	.985	.982	.979	.975	.971
5	1.00	.999	.999	.999	.999	.998	.998	.997	.997	.996
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.074	.067	.061	.055	.049	.044	.039	.035	.031	.028
1	.309	.289	.270	.251	.234	.217	.201	.186	.172	.159
2	.624	.601	.578	.555	.532	.509	.487	.464	.442	.420
3	.861	.847	.832	.816	.800	.783	.766	.748	.729	.710
4	.967	.962	.957	.951	.944	.937	.930	.922	.913	.904
5	.995	.995	.994	.992	.991	.989	.988	.986	.984	.981
6	1.00	1.00	1.00	.999	.999	.999	.999	.999	.999	.998

m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
	p									
0	.025	.022	.020	.017	.015	.013	.012	.010	.009	.008
1	.146	.134	.123	.112	.102	.093	.085	.077	.069	.062
2	.398	.377	.356	.336	.316	.297	.279	.261	.243	.227
3	.691	.671	.650	.629	.608	.587	.565	.544	.522	.500
4	.894	.883	.872	.860	.847	834	.820	.805	.790	.773
5	.978	.975	.972	.968	.964	.960	.955	.950	.944	.937
6	.998	.998	.997	.997	.996	.996	.995	.994	.993	.992

Table 7 D. Cumulative binomial probability distribution for $n = 8$

m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.923	.851	.784	.721	.663	.610	.560	.513	.470	.430
1	.997	.990	.978	.962	.943	.921	.897	.870	.842	.813
2	1.00	1.00	.999	.997	.994	.990	.985	.979	.971	.962
3	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.997	.995
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.394	.360	.328	.299	.272	.248	.225	.204	.185	.168
1	.783	.752	.721	.689	.657	.626	.594	.563	.533	.503
2	.951	.939	.926	.911	.895	.877	.859	.839	.819	.797
3	.993	.990	.987	.983	.979	.973	.967	.960	.952	.944
4	.999	.999	.999	.998	.997	.996	.995	.993	.992	.990
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.152	.137	.124	.111	.100	.090	.081	.072	.065	.058
1	.474	.446	.419	.392	.367	.343	.319	.297	.276	.255
2	.775	.751	.728	.703	.679	.653	.628	.603	.577	.552
3	.934	.924	.912	.900	.886	.872	.857	.841	.824	.806
4	.987	.984	.981	.977	.973	.968	.962	.956	.949	.942
5	.998	.998	.997	.997	.996	.995	.994	.992	.991	.989
6	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.051	.046	.041	.036	.032	.028	.025	.022	.019	.017
1	.236	.218	.201	.184	.169	.155	.141	.129	.117	.106
2	.526	.501	.476	.452	.428	.404	.381	.359	.337	.315
3	.787	.768	.748	.728	.706	.685	.663	.640	.617	.594
4	.934	.925	.915	.905	.894	.882	.869	.856	.841	.826
5	.987	.984	.981	.978	.975	.971	.966	.961	.956	.950
6	.998	.998	.998	.997	.996	.996	.995	.994	.993	.991
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999

m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.015	.013	.011	.010	.008	.007	.006	.005	.005	.004
1	.096	.087	.078	.070	.063	.057	.050	.045	.040	.035
2	.295	.275	.256	.238	.220	.203	.187	.172	.158	.145
3	.571	.547	.524	.500	.477	.454	.431	.408	.385	.363
4	.810	.794	.776	.758	.740	.720	.700	.680	.658	.637
5	.944	.937	.929	.921	.912	.902	.891	.880	.868	.855
6	.990	.988	.986	.984	.982	.979	.976	.973	.969	.965
7	.999	.999	.999	.999	.998	.998	.998	.997	.997	.996

Table 7 E. Cumulative binomial probability distribution for $n = 9$

m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.914	.834	.760	.693	.630	.573	.520	.472	.428	.387
1	.997	.987	.972	.952	.929	.902	.873	.842	.809	.775
2	1.00	.999	.998	.996	.992	.986	.979	.970	.960	.947
3	1.00	1.00	1.00	1.00	.999	.999	.998	.996	.994	.992
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.350	.316	.286	.257	.232	.208	.187	.168	.150	.134
1	.740	.705	.670	.634	.599	.565	.532	.499	.467	.436
2	.933	.917	.899	.880	.859	.837	.814	.790	.764	.738
3	.988	.984	.979	.973	.966	.958	.949	.938	.927	.914
4	.999	.998	.997	.996	.994	.993	.990	.988	.984	.980
5	1.00	1.00	1.00	1.00	.999	.999	.999	.998	.998	.997
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.120	.107	.095	.085	.075	.067	.059	.052	.046	.040
1	.407	.378	.351	.325	.300	.277	.255	.234	.214	.196
2	.711	.684	.657	.629	.601	.573	.545	.517	.490	.463
3	.901	.886	.870	.852	.834	.815	.795	.774	.752	.730
4	.976	.971	.965	.958	.951	.943	.934	.924	.913	.901
5	.996	.995	.994	.992	.990	.988	.985	.982	.979	.975
6	1.00	.999	.999	.999	.999	.998	.998	.997	.997	.996

	p									
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.035	.031	.027	.024	.021	.018	.016	.014	.012	.010
1	.179	.163	.148	.134	.121	.109	.098	.088	.079	.071
2	.436	.411	.385	.361	.337	.314	.292	.271	.251	.232
3	.706	.683	.658	.634	.609	.584	.558	.533	.508	.483
4	.888	.875	.860	.845	.828	.811	.793	.774	.754	.733
5	.970	.965	.960	.953	.946	.939	.930	.921	.911	.901
6	.995	.994	.992	.991	.989	.987	.984	.982	.978	.975
7	.999	.999	.999	.999	.999	.998	.998	.997	.997	.996

	p									
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.009	.007	.006	.005	.005	.004	.003	.003	.002	.002
1	.063	.056	.049	.044	.039	.034	.030	.026	.023	.020
2	.213	.196	.180	.164	.150	.136	.123	.111	.100	.090
3	.458	.433	.409	.385	.361	.339	.316	.295	.274	.254
4	.712	.690	.668	.645	.621	.598	.573	.549	.525	.500
5	.889	.877	.863	.849	.834	.818	.801	.784	.765	.746
6	.971	.967	.962	.956	.950	.944	.936	.928	.920	.910
7	.995	.995	.994	.992	.991	.989	.988	.985	.983	.980
8	1.00	1.00	.999	.999	.999	.999	.999	.999	.998	.998

Table 7 F. Cumulative binomial probability distribution for $n = 10$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.904	.817	.737	.665	.599	.539	.484	.434	.389	.349
1	.996	.984	.965	.942	.914	.882	.848	.812	.775	.736
2	1.00	.999	.997	.994	.988	.981	.972	.960	.946	.930
3	1.00	1.00	1.00	1.00	.999	.998	.996	.994	.991	.987
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.312	.279	.248	.221	.197	.175	.155	.137	.122	.107
1	.697	.658	.620	.582	.544	.508	.473	.439	.407	.376
2	.912	.891	.869	.845	.820	.794	.766	.737	.708	.678
3	.982	.976	.969	.960	.950	.939	.926	.912	.896	.879
4	.997	.996	.995	.993	.990	.987	.983	.979	.973	.967
5	1.00	1.00	.999	.999	.999	.998	.997	.996	.995	.994
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.095	.083	.073	.064	.056	.049	.043	.037	.033	.028
1	.346	.318	.292	.267	.244	.222	.202	.183	.166	.149
2	.647	.617	.586	.556	.526	.496	.466	.438	.410	.383
3	.861	.841	.821	.799	.776	.752	.727	.702	.676	.650
4	.960	.952	.943	.933	.922	.910	.896	.882	.866	.850
5	.992	.990	.987	.984	.980	.976	.971	.966	.960	.953
6	.999	.998	.998	.997	.996	.996	.994	.993	.991	.989
7	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.024	.021	.018	.016	.013	.012	.010	.008	.007	.006
1	.134	.121	.108	.096	.086	.076	.068	.060	.053	.046
2	.357	.331	.307	.284	.262	.241	.221	.202	.184	.167
3	.623	.596	.568	.541	.514	.487	.460	.434	.408	.382
4	.832	.813	.794	.773	.751	.729	.706	.682	.658	.633
5	.945	.936	.927	.916	.905	.893	.879	.865	.850	.834
6	.987	.984	.981	.978	.974	.969	.964	.959	.952	.945
7	.998	.997	.997	.996	.995	.994	.993	.991	.990	.988
8	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.999	.998
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.005	.004	.004	.003	.003	.002	.002	.001	.001	.001
1	.041	.036	.031	.027	.023	.020	.017	.015	.013	.011

2	.152	.137	.124	.111	.100	.089	.079	.070	.062	.055
3	.358	.333	.310	.288	.266	.245	.226	.207	.189	.172
4	.608	.582	.556	.530	.504	.478	.453	.427	.402	.377
5	.817	.798	.779	.759	.738	.717	.694	.671	.647	.623
6	.937	.929	.919	.909	.898	.886	.873	.859	.844	.828
7	.985	.983	.980	.976	.973	.968	.963	.958	.952	.945
8	.998	.997	.997	.996	.995	.995	.994	.992	.991	.989
9	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999

Table 7 G. Cumulative binomial probability distribution for $n = 11$

					p					
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.895	.801	.715	.638	.569	.506	.450	.400	.354	.314
1	.995	.980	.959	.931	.898	.862	.823	.782	.740	.697
2	1.00	.999	.996	.992	.985	.975	.963	.948	.931	.910
3	1.00	1.00	1.00	.999	.998	.997	.995	.991	.987	.981
4	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.997
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

					p					
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.278	.245	.216	.190	.167	.147	.129	.113	.098	.086
1	.655	.613	.571	.531	.492	.455	.419	.385	.353	.322
2	.888	.863	.837	.809	.779	.748	.716	.684	.651	.617
3	.974	.966	.956	.944	.931	.915	.899	.880	.860	.839
4	.996	.994	.991	.988	.984	.979	.973	.967	.959	.950
5	1.00	.999	.999	.998	.997	.996	.995	.993	.991	.988
6	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.998
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

					p					
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.075	.065	.056	.049	.042	.036	.031	.027	.023	.020
1	.294	.267	.242	.219	.197	.177	.159	.142	.127	.113
2	.584	.551	.519	.487	.455	.425	.395	.366	.339	.313
3	.816	.792	.767	.740	.713	.685	.657	.628	.599	.570
4	.939	.928	.915	.901	.885	.869	.851	.832	.811	.790
5	.985	.981	.977	.972	.966	.959	.951	.942	.933	.922
6	.997	.996	.995	.994	.992	.990	.988	.985	.982	.978
7	1.00	1.00	.999	.999	.999	.998	.998	.997	.997	.996
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.017	.014	.012	.010	.009	.007	.006	.005	.004	.004
1	.100	.089	.078	.069	.061	.053	.046	.040	.035	.030
2	.288	.264	.241	.220	.200	.181	.164	.148	.133	.119
3	.540	.511	.482	.454	.426	.398	.371	.345	.320	.296
4	.767	.744	.719	.694	.668	.642	.615	.588	.560	.533
5	.910	.897	.883	.868	.851	.834	.815	.796	.775	.753
6	.974	.969	.963	.957	.950	.942	.933	.923	.912	.901
7	.995	.993	.992	.990	.988	.985	.982	.979	.975	.971
8	.999	.999	.999	.998	.998	.997	.997	.996	.995	.994
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.003	.002	.002	.002	.001	.001	.001	.001	.001	.000
1	.026	.022	.019	.016	.014	.012	.010	.008	.007	.006
2	.106	.094	.084	.074	.065	.057	.050	.044	.038	.033
3	.273	.251	.230	.210	.191	.173	.157	.141	.127	.113
4	.505	.478	.451	.424	.397	.371	.346	.321	.297	.274
5	.731	.708	.683	.659	.633	.607	.581	.554	.527	.500
6	.888	.874	.859	.843	.826	.808	.789	.769	.748	.726
7	.966	.960	.954	.947	.939	.930	.921	.910	.899	.887
8	.993	.991	.990	.988	.985	.982	.979	.976	.972	.967
9	.999	.999	.999	.998	.998	.997	.997	.996	.995	.994
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 7 H. Cumulative binomial probability distribution for $n = 12$

m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.886	.785	.694	.613	.540	.476	.419	.368	.322	.282
1	.994	.977	.951	.919	.882	.840	.797	.751	.705	.659
2	1.00	.998	.995	.989	.980	.968	.953	.935	.913	.889
3	1.00	1.00	1.00	.999	.998	.996	.992	.988	.982	.974
4	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.997	.996
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.247	.216	.188	.164	.142	.123	.107	.092	.080	.069
1	.613	.569	.525	.483	.443	.405	.370	.336	.304	.275
2	.862	.833	.802	.770	.736	.701	.666	.630	.594	.558
3	.965	.954	.940	.925	.908	.889	.868	.845	.820	.795
4	.993	.991	.987	.982	.976	.969	.961	.951	.940	.927
5	.999	.999	.998	.997	.995	.994	.991	.988	.985	.981
6	1.00	1.00	1.00	1.00	.999	.999	.999	.998	.997	.996
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999

m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.059	.051	.043	.037	.032	.027	.023	.019	.016	.014
1	.248	.222	.199	.178	.158	.141	.125	.110	.097	.085
2	.523	.489	.455	.422	.391	.360	.331	.304	.278	.253
3	.767	.739	.710	.680	.649	.618	.586	.555	.524	.493
4	.913	.898	.881	.862	.842	.821	.798	.775	.750	.724
5	.976	.970	.963	.955	.946	.935	.924	.911	.897	.882
6	.995	.993	.991	.989	.986	.982	.978	.973	.968	.961
7	.999	.999	.998	.998	.997	.996	.995	.994	.992	.991
8	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998

m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.012	.010	.008	.007	.006	.005	.004	.003	.003	.002
1	.074	.065	.057	.049	.042	.037	.031	.027	.023	.020
2	.230	.208	.188	.169	.151	.135	.120	.107	.095	.083
3	.462	.432	.403	.374	.347	.320	.295	.270	.247	.225
4	.697	.669	.641	.612	.583	.554	.525	.496	.467	.438
5	.866	.848	.829	.809	.787	.765	.741	.717	.691	.665
6	.954	.946	.937	.927	.915	.903	.889	.875	.859	.842
7	.988	.986	.982	.979	.974	.970	.964	.958	.951	.943
8	.998	.997	.996	.996	.994	.993	.991	.990	.987	.985
9	1.00	1.00	1.00	.999	.999	.999	.999	.998	.998	.997
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.002	.001	.001	.001	.001	.001	.000	.000	.000	.000
1	.017	.014	.012	.010	.008	.007	.006	.005	.004	.003
2	.073	.064	.056	.049	.042	.036	.031	.027	.023	.019

3	.205	.185	.167	.150	.134	.120	.107	.094	.083	.073
4	.410	.383	.356	.330	.304	.280	.257	.235	.214	.194
5	.638	.611	.583	.555	.527	.499	.470	.442	.415	.387
6	.824	.804	.784	.762	.739	.716	.691	.666	.640	.613
7	.934	.924	.913	.901	.888	.874	.859	.842	.825	.806
8	.982	.978	.974	.970	.964	.959	.952	.944	.936	.927
9	.997	.996	.995	.994	.992	.990	.989	.986	.984	.981
10	1.00	.999	.999	.999	.999	.999	.998	.998	.997	.997
11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 7 I. Cumulative binomial probability distribution for $n = 13$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.878	.769	.673	.588	.513	.447	.389	.338	.293	.254
1	.993	.973	.944	.907	.865	.819	.770	.721	.671	.621
2	1.00	.998	.994	.986	.975	.961	.942	.920	.895	.866
3	1.00	1.00	1.00	.999	.997	.994	.990	.984	.976	.966
4	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.996	.994
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.220	.190	.164	.141	.121	.104	.089	.076	.065	.055
1	.573	.526	.481	.439	.398	.360	.325	.292	.262	.234
2	.835	.802	.766	.730	.692	.654	.615	.577	.539	.502
3	.954	.939	.922	.903	.882	.859	.833	.806	.777	.747
4	.990	.986	.981	.974	.966	.956	.945	.932	.917	.901
5	.998	.998	.996	.995	.992	.990	.986	.982	.976	.970
6	1.00	1.00	.999	.999	.999	.998	.997	.996	.995	.993
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.047	.040	.033	.028	.024	.020	.017	.014	.012	.010
1	.208	.185	.163	.144	.127	.111	.097	.085	.074	.064
2	.465	.430	.396	.364	.333	.303	.275	.249	.225	.202
3	.716	.684	.651	.618	.584	.551	.517	.485	.452	.421
4	.883	.863	.841	.818	.794	.768	.741	.713	.684	.654
5	.962	.954	.944	.932	.920	.906	.890	.873	.855	.835
6	.991	.988	.985	.981	.976	.970	.963	.956	.947	.938

7	.998	.998	.997	.996	.994	.993	.991	.988	.985	.982
8	1.00	1.00	.999	.999	.999	.999	.998	.998	.997	.996
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999

					p					
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.008	.007	.005	.005	.004	.003	.002	.002	.002	.001
1	.055	.047	.041	.035	.030	.025	.021	.018	.015	.013
2	.181	.162	.144	.128	.113	.100	.088	.077	.067	.058
3	.390	.360	.332	.304	.278	.254	.230	.208	.188	.169
4	.624	.593	.562	.531	.501	.470	.440	.410	.381	.353
5	.813	.791	.767	.742	.716	.689	.661	.633	.604	.574
6	.927	.915	.901	.887	.871	.853	.835	.815	.794	.771
7	.978	.973	.967	.961	.954	.946	.936	.926	.915	.902
8	.995	.993	.992	.990	.987	.985	.981	.977	.973	.968
9	.999	.999	.998	.998	.997	.997	.996	.995	.994	.992
10	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999

					p					
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.001	.001	.001	.001	.000	.000	.000	.000	.000	.000
1	.011	.009	.007	.006	.005	.004	.003	.003	.002	.002
2	.050	.043	.037	.032	.027	.023	.019	.016	.014	.011
3	.151	.134	.119	.106	.093	.081	.071	.062	.054	.046
4	.326	.300	.275	.251	.228	.206	.186	.167	.150	.133
5	.545	.515	.485	.456	.427	.398	.370	.343	.316	.291
6	.748	.723	.697	.671	.644	.616	.587	.558	.529	.500
7	.889	.874	.857	.840	.821	.801	.780	.758	.734	.709
8	.962	.955	.948	.940	.930	.920	.908	.896	.882	.867
9	.990	.988	.986	.983	.980	.976	.971	.966	.960	.954
10	.998	.998	.997	.997	.996	.995	.994	.992	.991	.989
11	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.999	.998

Table 7 J. Cumulative binomial probability distribution for $n = 14$

					p					
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.869	.754	.653	.565	.488	.421	.362	.311	.267	.229
1	.992	.969	.936	.894	.847	.796	.744	.690	.637	.585
2	1.00	.998	.992	.983	.970	.952	.930	.904	.874	.842
3	1.00	1.00	.999	.998	.996	.992	.986	.979	.969	.956

4	1.00	1.00	1.00	1.00	1.00	.999	.998	.996	.994	.991
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

					p					
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.196	.167	.142	.121	.103	.087	.074	.062	.052	.044
1	.534	.486	.440	.397	.357	.319	.285	.253	.224	.198
2	.806	.768	.729	.689	.648	.607	.566	.526	.486	.448
3	.941	.923	.902	.879	.853	.826	.796	.765	.732	.698
4	.986	.980	.973	.964	.953	.941	.926	.909	.891	.870
5	.998	.996	.994	.992	.988	.984	.979	.973	.965	.956
6	1.00	.999	.999	.999	.998	.997	.995	.994	.991	.988
7	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.998	.998

					p					
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.037	.031	.026	.021	.018	.015	.012	.010	.008	.007
1	.174	.153	.133	.116	.101	.087	.075	.065	.056	.047
2	.411	.376	.343	.311	.281	.253	.227	.203	.181	.161
3	.663	.628	.592	.557	.521	.486	.452	.419	.386	.355
4	.848	.824	.798	.770	.742	.712	.681	.649	.617	.584
5	.946	.934	.920	.905	.888	.870	.850	.828	.805	.781
6	.985	.980	.975	.969	.962	.953	.944	.933	.920	.907
7	.997	.995	.994	.992	.990	.987	.983	.979	.974	.969
8	.999	.999	.999	.998	.998	.997	.996	.995	.994	.992
9	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.998

					p					
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.006	.005	.004	.003	.002	.002	.002	.001	.001	.001
1	.040	.034	.029	.024	.021	.017	.014	.012	.010	.008
2	.142	.125	.110	.096	.084	.073	.063	.054	.047	.040
3	.325	.297	.270	.244	.220	.198	.177	.158	.141	.124
4	.551	.519	.486	.454	.423	.392	.362	.333	.306	.279
5	.755	.728	.699	.670	.641	.610	.579	.548	.517	.486
6	.892	.875	.857	.837	.816	.794	.770	.746	.720	.692
7	.962	.954	.946	.936	.925	.912	.899	.884	.868	.850
8	.990	.987	.984	.980	.976	.971	.965	.958	.950	.942
9	.998	.997	.996	.995	.994	.992	.991	.988	.986	.982

10	1.00	1.00	.999	.999	.999	.999	.998	.998	.997	.996
11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.007	.005	.004	.004	.003	.002	.002	.001	.001	.001
2	.034	.029	.024	.020	.017	.014	.012	.010	.008	.006
3	.110	.096	.084	.073	.063	.054	.047	.040	.034	.029
4	.254	.230	.208	.187	.167	.149	.132	.117	.103	.090
5	.455	.425	.395	.366	.337	.310	.284	.259	.235	.212
6	.664	.636	.606	.576	.546	.516	.485	.455	.425	.395
7	.831	.810	.789	.766	.741	.716	.689	.662	.634	.605
8	.932	.921	.909	.896	.881	.865	.848	.829	.809	.788
9	.979	.975	.970	.964	.957	.950	.942	.932	.922	.910
10	.995	.994	.992	.991	.989	.986	.983	.980	.976	.971
11	.999	.999	.999	.998	.998	.997	.997	.996	.995	.994
12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999
13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 7 K. Cumulative binomial probability distribution for $n = 15$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.860	.739	.633	.542	.463	.395	.337	.286	.243	.206
1	.990	.965	.927	.881	.829	.774	.717	.660	.604	.549
2	1.00	.997	.991	.980	.964	.943	.917	.887	.853	.816
3	1.00	1.00	.999	.998	.995	.990	.982	.973	.960	.944
4	1.00	1.00	1.00	1.00	.999	.999	.997	.995	.992	.987
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.174	.147	.124	.104	.087	.073	.061	.051	.042	.035
1	.497	.448	.401	.358	.319	.282	.249	.219	.192	.167
2	.776	.735	.692	.648	.604	.561	.518	.477	.436	.398
3	.926	.904	.880	.852	.823	.791	.757	.722	.685	.648
4	.981	.974	.964	.952	.938	.922	.904	.883	.861	.836
5	.996	.994	.992	.988	.983	.977	.970	.961	.951	.939
6	.999	.999	.998	.998	.996	.995	.993	.990	.986	.982
7	1.00	1.00	1.00	1.00	.999	.999	.999	.998	.997	.996
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999

m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.029	.024	.020	.016	.013	.011	.009	.007	.006	.005
1	.145	.126	.109	.094	.080	.069	.058	.050	.042	.035
2	.361	.327	.294	.264	.236	.210	.186	.165	.145	.127
3	.610	.573	.535	.498	.461	.426	.391	.358	.327	.297
4	.809	.781	.750	.719	.686	.653	.619	.585	.550	.515
5	.925	.910	.892	.873	.852	.829	.804	.778	.750	.722
6	.977	.970	.963	.954	.943	.932	.918	.903	.887	.869
7	.994	.992	.990	.987	.983	.978	.973	.966	.959	.950
8	.999	.998	.998	.997	.996	.994	.993	.991	.988	.985
9	1.00	1.00	1.00	.999	.999	.999	.998	.998	.997	.996
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999

m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.004	.003	.002	.002	.002	.001	.001	.001	.001	.000
1	.030	.025	.021	.017	.014	.012	.010	.008	.006	.005
2	.111	.096	.083	.072	.062	.053	.045	.038	.032	.027
3	.269	.242	.217	.194	.173	.153	.135	.119	.104	.091
4	.481	.448	.415	.383	.352	.322	.294	.267	.241	.217
5	.692	.661	.629	.597	.564	.532	.499	.467	.435	.403
6	.849	.828	.805	.781	.755	.728	.700	.671	.641	.610
7	.940	.929	.916	.902	.887	.870	.851	.831	.810	.787
8	.981	.976	.971	.965	.958	.950	.940	.930	.918	.905
9	.995	.994	.992	.990	.988	.985	.981	.977	.972	.966
10	.999	.999	.998	.998	.997	.996	.995	.994	.993	.991
11	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998

m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.004	.003	.003	.002	.002	.001	.001	.001	.001	.000
2	.023	.019	.016	.013	.011	.009	.007	.006	.005	.004
3	.078	.068	.058	.050	.042	.036	.030	.025	.021	.018
4	.195	.174	.155	.137	.120	.106	.092	.080	.069	.059
5	.373	.343	.314	.287	.261	.236	.212	.190	.170	.151
6	.579	.547	.515	.484	.452	.421	.391	.361	.332	.304
7	.763	.737	.710	.682	.654	.624	.593	.563	.531	.500
8	.890	.875	.857	.838	.818	.797	.773	.749	.723	.696
9	.960	.952	.944	.934	.923	.911	.898	.883	.867	.849

10	.988	.986	.983	.979	.975	.969	.964	.957	.949	.941
11	.998	.997	.996	.995	.994	.992	.990	.988	.985	.982
12	1.00	1.00	.999	.999	.999	.999	.998	.998	.997	.996
13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1000

Table 7 L. Cumulative binomial probability distribution for $n = 16$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.851	.724	.614	.520	.440	.372	.313	.263	.221	.185
1	.989	.960	.918	.867	.811	.751	.690	.630	.571	.515
2	.999	.996	.989	.976	.957	.933	.903	.869	.831	.789
3	1.00	1.00	.999	.997	.993	.987	.978	.966	.950	.932
4	1.00	1.00	1.00	1.00	.999	.998	.996	.993	.989	.983
5	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.997
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.155	.129	.108	.090	.074	.061	.051	.042	.034	.028
1	.461	.412	.365	.323	.284	.249	.217	.189	.163	.141
2	.745	.700	.654	.607	.561	.516	.472	.430	.390	.352
3	.909	.884	.855	.824	.790	.754	.716	.678	.638	.598
4	.975	.965	.953	.938	.921	.901	.879	.854	.827	.798
5	.995	.992	.988	.983	.976	.968	.959	.947	.934	.918
6	.999	.998	.998	.996	.994	.992	.989	.985	.980	.973
7	1.00	1.00	1.00	.999	.999	.998	.998	.996	.995	.993
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.023	.019	.015	.012	.010	.008	.007	.005	.004	.003
1	.121	.103	.088	.075	.063	.054	.045	.038	.031	.026
2	.316	.283	.252	.223	.197	.173	.152	.132	.115	.099
3	.558	.519	.480	.442	.405	.370	.336	.304	.274	.246
4	.767	.735	.701	.666	.630	.594	.558	.521	.485	.450
5	.901	.881	.860	.836	.810	.783	.754	.724	.692	.660
6	.966	.957	.946	.934	.920	.905	.888	.868	.847	.825
7	.990	.987	.983	.979	.973	.966	.958	.949	.938	.926
8	.998	.997	.996	.994	.993	.990	.987	.984	.979	.974
9	1.00	.999	.999	.999	.998	.998	.997	.996	.995	.993
10	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.998
11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.003	.002	.002	.001	.001	.001	.001	.000	.000	.000
1	.022	.018	.015	.012	.010	.008	.006	.005	.004	.003
2	.086	.073	.063	.053	.045	.038	.032	.027	.022	.018
3	.220	.195	.173	.153	.134	.117	.102	.088	.076	.065
4	.415	.382	.350	.319	.289	.261	.235	.211	.188	.167
5	.626	.593	.558	.524	.490	.456	.423	.391	.359	.329
6	.800	.774	.747	.718	.688	.657	.625	.593	.560	.527
7	.912	.897	.880	.861	.841	.819	.795	.770	.744	.716
8	.968	.961	.953	.944	.933	.921	.907	.892	.876	.858
9	.991	.988	.985	.982	.977	.972	.966	.959	.951	.942
10	.998	.997	.996	.995	.994	.992	.990	.987	.984	.981
11	1.00	.999	.999	.999	.999	.998	.998	.997	.996	.995
12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999

m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.003	.002	.002	.001	.001	.001	.001	.000	.000	.000
2	.015	.012	.010	.008	.007	.005	.004	.003	.003	.002
3	.056	.047	.040	.034	.028	.023	.019	.016	.013	.011
4	.147	.129	.113	.098	.085	.074	.063	.054	.046	.038
5	.300	.272	.246	.221	.198	.176	.156	.137	.120	.105
6	.494	.461	.429	.397	.366	.336	.307	.279	.252	.227
7	.687	.657	.626	.595	.563	.531	.498	.466	.434	.402
8	.838	.817	.794	.770	.744	.717	.689	.660	.629	.598
9	.931	.920	.906	.892	.876	.858	.839	.819	.796	.773
10	.977	.972	.966	.959	.951	.943	.933	.921	.909	.895
11	.994	.992	.990	.988	.985	.982	.978	.973	.968	.962
12	.999	.998	.998	.997	.997	.996	.994	.993	.991	.989
13	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998	.998

Table 7 M. Cumulative binomial probability distribution for $n = 17$

m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.843	.709	.596	.500	.418	.349	.291	.242	.201	.167
1	.988	.955	.909	.853	.792	.728	.664	.601	.540	.482
2	.999	.996	.987	.971	.950	.922	.888	.850	.807	.762
3	1.00	1.00	.999	.996	.991	.984	.973	.958	.940	.917
4	1.00	1.00	1.00	1.00	.999	.997	.995	.991	.985	.978

5	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.997	.995
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.138	.114	.094	.077	.063	.052	.042	.034	.028	.023
1	.428	.378	.332	.290	.252	.219	.189	.162	.139	.118
2	.714	.665	.616	.568	.520	.473	.429	.387	.347	.310
3	.891	.862	.829	.793	.756	.716	.675	.633	.591	.549
4	.968	.955	.940	.922	.901	.878	.851	.822	.791	.758
5	.993	.989	.983	.977	.968	.958	.945	.931	.914	.894
6	.999	.998	.996	.994	.992	.988	.984	.978	.971	.962
7	1.00	1.00	.999	.999	.998	.997	.996	.994	.992	.989
8	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.997

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.018	.015	.012	.009	.008	.006	.005	.004	.003	.002
1	.100	.085	.071	.060	.050	.042	.035	.029	.024	.019
2	.275	.243	.214	.188	.164	.142	.123	.106	.091	.077
3	.507	.467	.427	.389	.353	.319	.286	.256	.228	.202
4	.723	.687	.650	.612	.574	.536	.498	.460	.424	.389
5	.873	.849	.823	.795	.765	.734	.701	.667	.632	.597
6	.952	.940	.926	.911	.893	.873	.852	.828	.802	.775
7	.985	.981	.975	.968	.960	.950	.939	.926	.912	.895
8	.996	.995	.993	.991	.988	.984	.979	.974	.967	.960
9	.999	.999	.998	.998	.997	.996	.994	.992	.990	.987
10	1.00	1.00	1.00	1.00	.999	.999	.999	.998	.998	.997
11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999

	p									
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.002	.001	.001	.001	.001	.001	.000	.000	.000	.000
1	.016	.013	.010	.008	.007	.005	.004	.003	.003	.002
2	.066	.056	.047	.039	.033	.027	.022	.018	.015	.012
3	.178	.156	.137	.119	.103	.089	.076	.065	.055	.046
4	.355	.322	.291	.262	.235	.209	.186	.164	.144	.126
5	.561	.525	.489	.454	.420	.386	.353	.322	.292	.264
6	.746	.716	.685	.652	.619	.585	.551	.516	.482	.448
7	.877	.857	.836	.812	.787	.761	.732	.703	.672	.641
8	.951	.941	.929	.916	.901	.884	.866	.846	.824	.801
9	.984	.980	.975	.969	.962	.954	.944	.934	.922	.908

10	.996	.994	.993	.991	.988	.985	.981	.977	.971	.965
11	.999	.999	.998	.998	.997	.996	.995	.993	.992	.989
12	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998	.997
13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

					p					
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.002	.001	.001	.001	.001	.000	.000	.000	.000	.000
2	.010	.008	.006	.005	.004	.003	.003	.002	.002	.001
3	.039	.033	.027	.022	.018	.015	.012	.010	.008	.006
4	.110	.095	.082	.070	.060	.050	.043	.036	.030	.025
5	.237	.212	.189	.167	.147	.129	.112	.097	.084	.072
6	.414	.382	.350	.320	.290	.262	.236	.211	.188	.166
7	.608	.575	.541	.508	.474	.441	.408	.376	.345	.315
8	.776	.750	.722	.693	.663	.631	.599	.566	.533	.500
9	.893	.876	.858	.838	.817	.793	.769	.742	.715	.685
10	.958	.950	.940	.930	.917	.904	.889	.872	.854	.834
11	.987	.984	.980	.975	.970	.964	.957	.948	.939	.928
12	.997	.996	.995	.993	.991	.989	.987	.984	.980	.975
13	.999	.999	.999	.999	.998	.998	.997	.996	.995	.994
14	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999
15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 7 N. Cumulative binomial probability distribution for $n = 18$

					p					
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.835	.695	.578	.480	.397	.328	.271	.223	.183	.150
1	.986	.950	.900	.839	.774	.706	.638	.572	.509	.450
2	.999	.995	.984	.967	.942	.910	.873	.830	.783	.734
3	1.00	1.00	.998	.995	.989	.980	.967	.949	.928	.902
4	1.00	1.00	1.00	.999	.998	.997	.993	.988	.981	.972
5	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.996	.994
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.123	.100	.082	.066	.054	.043	.035	.028	.023	.018
1	.396	.346	.301	.260	.224	.192	.164	.139	.118	.099
2	.683	.631	.579	.529	.480	.433	.388	.346	.307	.271
3	.872	.838	.801	.762	.720	.677	.633	.589	.545	.501
4	.959	.944	.926	.904	.879	.852	.821	.788	.753	.716
5	.990	.985	.978	.969	.958	.945	.929	.911	.890	.867
6	.998	.997	.995	.992	.988	.983	.977	.969	.960	.949
7	1.00	.999	.999	.998	.997	.996	.994	.991	.988	.984
8	1.00	1.00	1.00	1.00	.999	.999	.999	.998	.997	.996
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.014	.011	.009	.007	.006	.004	.003	.003	.002	.002
1	.083	.069	.058	.048	.039	.032	.027	.022	.018	.014
2	.238	.208	.181	.157	.135	.116	.099	.084	.071	.060
3	.459	.418	.378	.341	.306	.273	.242	.214	.188	.165
4	.678	.639	.599	.559	.519	.479	.441	.403	.367	.333
5	.841	.813	.783	.751	.717	.682	.646	.609	.572	.534
6	.935	.920	.903	.883	.861	.837	.811	.783	.753	.722
7	.978	.972	.964	.954	.943	.930	.915	.899	.880	.859
8	.994	.992	.989	.985	.981	.975	.968	.961	.951	.940
9	.999	.998	.997	.996	.995	.993	.990	.987	.984	.979
10	1.00	1.00	.999	.999	.999	.998	.998	.997	.995	.994
11	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999

	p									
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.001	.001	.001	.001	.000	.000	.000	.000	.000	.000
1	.011	.009	.007	.006	.005	.004	.003	.002	.002	.001
2	.050	.042	.035	.029	.024	.019	.016	.013	.010	.008
3	.143	.124	.107	.092	.078	.066	.056	.047	.039	.033
4	.300	.269	.240	.213	.189	.166	.145	.126	.109	.094
5	.497	.460	.424	.389	.355	.322	.291	.262	.235	.209
6	.689	.655	.620	.585	.549	.513	.478	.442	.408	.374
7	.837	.812	.786	.758	.728	.697	.665	.632	.598	.563
8	.928	.914	.898	.880	.861	.840	.816	.792	.765	.737
9	.974	.967	.959	.951	.940	.929	.915	.900	.884	.865
10	.992	.990	.987	.983	.979	.974	.967	.960	.952	.942

11	.998	.997	.996	.995	.994	.992	.990	.987	.984	.980
12	1.00	.999	.999	.999	.999	.998	.997	.997	.996	.994
13	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999

	p									
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.001	.001	.001	.000	.000	.000	.000	.000	.000	.000
2	.007	.005	.004	.003	.003	.002	.001	.001	.001	.001
3	.027	.022	.018	.015	.012	.010	.008	.006	.005	.004
4	.081	.069	.058	.049	.041	.034	.028	.023	.019	.015
5	.185	.163	.143	.124	.108	.093	.079	.068	.057	.048
6	.342	.311	.281	.252	.226	.201	.178	.156	.137	.119
7	.529	.494	.459	.425	.391	.359	.327	.297	.268	.240
8	.707	.676	.644	.611	.578	.544	.509	.475	.441	.407
9	.845	.823	.800	.774	.747	.719	.689	.658	.626	.593
10	.931	.919	.905	.889	.872	.853	.832	.810	.786	.760
11	.975	.969	.963	.955	.946	.936	.925	.912	.897	.881
12	.993	.991	.988	.985	.982	.977	.972	.967	.960	.952
13	.998	.998	.997	.996	.995	.994	.992	.990	.988	.985
14	1.00	1.00	.999	.999	.999	.999	.998	.998	.997	.996
15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999

Table 7 O. Cumulative binomial probability distribution for $n = 19$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.826	.681	.561	.460	.377	.309	.252	.205	.167	.135
1	.985	.945	.890	.825	.755	.683	.612	.544	.480	.420
2	.999	.994	.982	.962	.933	.898	.856	.809	.759	.705
3	1.00	1.00	.998	.994	.987	.976	.960	.940	.915	.885
4	1.00	1.00	1.00	.999	.998	.996	.991	.985	.977	.965
5	1.00	1.00	1.00	1.00	1.00	.999	.999	.997	.995	.991
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.109	.088	.071	.057	.046	.036	.029	.023	.018	.014
1	.366	.317	.272	.233	.198	.168	.142	.119	.100	.083
2	.651	.597	.543	.491	.441	.394	.350	.309	.271	.237
3	.851	.813	.772	.729	.684	.638	.591	.545	.499	.455
4	.950	.931	.910	.884	.856	.824	.789	.752	.714	.673
5	.986	.980	.971	.960	.946	.930	.911	.889	.864	.837

6	.997	.995	.992	.989	.984	.977	.969	.959	.947	.932
7	.999	.999	.998	.997	.996	.994	.991	.987	.983	.977
8	1.00	1.00	1.00	.999	.999	.999	.998	.997	.995	.993
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.011	.009	.007	.005	.004	.003	.003	.002	.001	.001
1	.069	.057	.047	.038	.031	.025	.020	.016	.013	.010
2	.206	.178	.153	.131	.111	.094	.079	.067	.056	.046
3	.412	.372	.333	.297	.263	.232	.204	.178	.154	.133
4	.632	.590	.548	.506	.465	.426	.387	.350	.315	.282
5	.807	.775	.741	.705	.668	.630	.591	.552	.513	.474
6	.916	.897	.875	.851	.825	.797	.766	.734	.701	.666
7	.969	.960	.950	.937	.923	.906	.887	.866	.843	.818
8	.991	.987	.983	.978	.971	.963	.954	.943	.931	.916
9	.998	.997	.995	.993	.991	.988	.984	.980	.974	.967
10	1.00	.999	.999	.998	.998	.997	.996	.994	.992	.989
11	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.998	.997
12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000
1	.008	.007	.005	.004	.003	.002	.002	.001	.001	.001
2	.038	.031	.026	.021	.017	.014	.011	.009	.007	.005
3	.114	.098	.083	.070	.059	.049	.041	.034	.028	.023
4	.251	.223	.196	.172	.150	.130	.112	.096	.082	.070
5	.436	.399	.363	.329	.297	.266	.237	.210	.186	.163
6	.629	.593	.556	.518	.481	.445	.409	.374	.340	.308
7	.791	.762	.731	.699	.666	.631	.596	.560	.524	.488
8	.900	.881	.861	.839	.815	.788	.760	.731	.700	.667
9	.959	.950	.939	.927	.913	.897	.879	.859	.837	.814
10	.986	.982	.978	.972	.965	.957	.948	.938	.925	.912
11	.996	.995	.993	.991	.989	.985	.982	.977	.971	.965
12	.999	.999	.998	.998	.997	.996	.995	.993	.991	.988
13	1.00	1.00	1.00	1.00	.999	.999	.999	.998	.998	.997
14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
2	.004	.003	.003	.002	.002	.001	.001	.001	.000	.000
3	.019	.015	.012	.010	.008	.006	.005	.004	.003	.002
4	.059	.049	.041	.034	.028	.023	.019	.015	.012	.010
5	.142	.123	.106	.091	.078	.066	.055	.046	.039	.032
6	.277	.248	.221	.196	.173	.151	.132	.114	.098	.084
7	.452	.417	.382	.349	.317	.286	.257	.229	.204	.180
8	.634	.600	.565	.529	.494	.459	.424	.390	.356	.324
9	.789	.762	.733	.703	.671	.638	.605	.570	.535	.500
10	.896	.879	.860	.839	.816	.791	.765	.737	.707	.676
11	.957	.948	.938	.926	.913	.898	.881	.863	.842	.820
12	.985	.982	.977	.972	.966	.958	.950	.940	.929	.916
13	.996	.995	.993	.991	.989	.986	.983	.979	.974	.968
14	.999	.999	.998	.998	.997	.996	.995	.994	.992	.990
15	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998	.998
16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 7 P. Cumulative binomial probability distribution for $n = 20$

	p									
m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0	.818	.668	.544	.442	.358	.290	.234	.189	.152	.122
1	.983	.940	.880	.810	.736	.660	.587	.517	.452	.392
2	.999	.993	.979	.956	.925	.885	.839	.788	.733	.677
3	1.00	.999	.997	.993	.984	.971	.953	.929	.901	.867
4	1.00	1.00	1.00	.999	.997	.994	.989	.982	.971	.957
5	1.00	1.00	1.00	1.00	1.00	.999	.998	.996	.993	.989
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998

	p									
m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
0	.097	.078	.062	.049	.039	.031	.024	.019	.015	.012
1	.338	.289	.246	.208	.176	.147	.123	.102	.084	.069
2	.620	.563	.508	.455	.405	.358	.315	.275	.239	.206
3	.829	.787	.743	.696	.648	.599	.550	.503	.456	.411
4	.939	.917	.892	.863	.830	.794	.756	.715	.673	.630
5	.982	.974	.963	.949	.933	.913	.890	.864	.836	.804
6	.996	.993	.990	.985	.978	.970	.959	.946	.931	.913
7	.999	.999	.998	.996	.994	.991	.987	.982	.976	.968

8	1.00	1.00	1.00	.999	.999	.998	.997	.995	.993	.990
9	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.997
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999

	p									
m	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30
0	.009	.007	.005	.004	.003	.002	.002	.001	.001	.001
1	.057	.046	.037	.030	.024	.019	.016	.012	.010	.008
2	.177	.151	.128	.109	.091	.076	.064	.053	.043	.035
3	.369	.329	.292	.257	.225	.196	.170	.147	.126	.107
4	.586	.542	.499	.456	.415	.375	.338	.302	.269	.238
5	.770	.734	.696	.657	.617	.577	.536	.495	.455	.416
6	.893	.870	.844	.816	.786	.753	.719	.683	.646	.608
7	.958	.946	.933	.917	.898	.878	.855	.829	.802	.772
8	.986	.981	.975	.968	.959	.948	.936	.922	.905	.887
9	.996	.995	.992	.990	.986	.982	.976	.970	.962	.952
10	.999	.999	.998	.997	.996	.995	.993	.990	.987	.983
11	1.00	1.00	1.00	.999	.999	.999	.998	.997	.996	.995
12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999
13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
0	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.006	.005	.004	.003	.002	.002	.001	.001	.001	.001
2	.029	.023	.019	.015	.012	.010	.008	.006	.005	.004
3	.091	.077	.064	.054	.044	.037	.030	.024	.020	.016
4	.209	.183	.159	.137	.118	.101	.086	.073	.061	.051
5	.379	.343	.308	.276	.245	.217	.191	.167	.145	.126
6	.569	.531	.492	.454	.417	.380	.345	.312	.280	.250
7	.741	.708	.673	.638	.601	.564	.527	.489	.452	.416
8	.866	.843	.818	.791	.762	.732	.700	.666	.631	.596
9	.941	.928	.913	.897	.878	.858	.835	.810	.784	.755
10	.978	.972	.965	.957	.947	.935	.922	.908	.891	.872
11	.993	.991	.988	.985	.980	.975	.969	.962	.953	.943
12	.998	.998	.997	.995	.994	.992	.990	.987	.983	.979
13	1.00	.999	.999	.999	.998	.998	.997	.996	.995	.994
14	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.998
15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	p									
m	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
2	.003	.002	.002	.001	.001	.001	.001	.000	.000	.000
3	.013	.010	.008	.006	.005	.004	.003	.002	.002	.001
4	.042	.035	.029	.023	.019	.015	.012	.010	.008	.006
5	.108	.092	.078	.066	.055	.046	.038	.031	.026	.021
6	.222	.196	.172	.150	.130	.112	.096	.081	.069	.058
7	.380	.346	.313	.282	.252	.224	.198	.174	.152	.132
8	.559	.523	.486	.450	.414	.379	.345	.313	.281	.252
9	.725	.694	.661	.626	.591	.556	.520	.483	.447	.412
10	.852	.830	.805	.779	.751	.721	.690	.657	.623	.588
11	.932	.919	.904	.888	.869	.849	.827	.802	.776	.748
12	.974	.968	.960	.952	.942	.931	.918	.903	.887	.868
13	.992	.989	.986	.983	.979	.973	.967	.960	.952	.942
14	.998	.997	.996	.995	.994	.992	.989	.987	.983	.979
15	1.00	.999	.999	.999	.998	.998	.997	.996	.995	.994
16	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999
17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 8. Quantities of Kolmogorov-Smirnov test statistic

One-sided test					
$p =$	0.90	0.95	0.975	0.99	0.995
Two-sided test					
$p =$	0.80	0.90	0.95	0.98	0.99
$n = 1$	0.900	0.950	0.975	0.990	0.995
2	0.684	0.776	0.842	0.900	0.929
3	0.565	0.636	0.708	0.785	0.829
4	0.493	0.565	0.624	0.689	0.734
5	0.447	0.509	0.563	0.627	0.669
6	0.410	0.468	0.519	0.577	0.617
7	0.381	0.436	0.483	0.538	0.576
8	0.358	0.410	0.454	0.507	0.542
9	0.339	0.387	0.430	0.480	0.513
10	0.323	0.369	0.409	0.457	0.489
11	0.308	0.352	0.391	0.437	0.468
12	0.296	0.338	0.375	0.419	0.449
13	0.285	0.325	0.361	0.404	0.432

14	0.275	0.314	0.349	0.390	0.418
15	0.266	0.304	0.338	0.377	0.404
16	0.258	0.295	0.327	0.366	0.392
17	0.250	0.286	0.318	0.355	0.381
18	0.244	0.279	0.309	0.346	0.371
19	0.237	0.271	0.301	0.337	0.361
20	0.232	0.265	0.294	0.329	0.352
21	0.226	0.259	0.287	0.321	0.344
22	0.221	0.253	0.281	0.314	0.337
23	0.216	0.247	0.275	0.307	0.330
24	0.212	0.242	0.269	0.301	0.323
25	0.208	0.238	0.264	0.295	0.317
26	0.204	0.233	0.259	0.290	0.311
27	0.200	0.229	0.254	0.284	0.305
28	0.197	0.225	0.250	0.279	0.300
29	0.193	0.221	0.246	0.275	0.295
30	0.190	0.218	0.242	0.270	0.290
31	0.187	0.214	0.238	0.266	0.285
32	0.184	0.211	0.234	0.262	0.281
33	0.182	0.208	0.231	0.258	0.277
34	0.179	0.205	0.227	0.254	0.273
35	0.177	0.202	0.224	0.251	0.269
36	0.174	0.199	0.221	0.247	0.265
37	0.172	0.196	0.218	0.244	0.262
38	0.170	0.194	0.215	0.241	0.258
39	0.168	0.191	0.213	0.238	0.255
40	0.165	0.189	0.210	0.235	0.252
Approximation for $n > 40$	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Bibliography

S. Chandra, Computer Applications in Physics with FORTRAN, BASIC and C (2006) Second Edition, Narosa Publishing House Pvt. Ltd., New Delhi

S. Chandra, Applications of numerical techniques with C (2006), Narosa Publishing House Pvt. Ltd., New Delhi

K.K. Chawla, Vijay Gupta, B.K. Sharma, Operations Research (2008) XII Edition, Kalyani Publishers, Ludhiana

W.W. Daniel, Biostatistics - Basic Concepts and Methodology for the Health Sciences, Ninth Edition (2010) Wiley India Pvt. Ltd., New Delhi

C. Dawson, Practical Research Methods (2002) UBS Publishers' Distributors, New Delhi

C.R. Kothari, Research Methodology - Methods and Techniques (2004) Second Edition, New Age International Publishers, New Delhi

R. Kumar, Research Methodology - A step-by-step Guide for Beginners, Second Edition (2005) Pearson Education, Singapore

J.B. Scarborough, Numerical Mathematical Analysis, sixth edition, Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi

S.D. Sharma, Operations Research (2005) XV Edition, Kedar Nath Ram Nath & Co., Meerut

K. Swarup, P.K. Gupta, M. Mohan, Operations Research (2007) XIII Edition, Sultan Chand & Sons, New Delhi

http://www.ihmctan.edu/PDF/notes/Research_Methodology.pdf

Index