



P.O. Box 342-01000 Thika

Email: Info@mku.ac.ke

Web: www.mku.ac.ke

DEPARTMENT OF MANAGEMENT

COURSE CODE: BBM 223

COURSE TITLE: INTRODUCTION TO BUSINESS

STATISTICS

Instructional Material for BBM- Distance Learning

Course content

Pre-requisites: BBM 112 and BBM 212

Purpose: To introduce students to business statistics necessary for data summarization and presentation

Expected Learning Outcomes of the Course:

By the end of the course unit the learners should be able to:-

- i) Organize and present data using various methods
- ii) Interpret summarized data

Course Content:

Data collection; Organization and Presentation of Data; Random variables: Discrete and continuous random variables, Their distribution such as binomial, Poisson, normal, and their business applications

Teaching / Learning Methodologies: Lectures and tutorials; group discussion; demonstration; Individual assignment; Case studies

Instructional Materials and Equipment: Projector; test books; design catalogues; computer laboratory; design software; simulators

Course Assessment

Examination - 70%; Continuous Assessment Test (CATS) - 20%; Assignments - 10%; Total - 100%

Recommended Text Books

- i) Azel (2006); *Complete Business Statistics*; Tata Mcgraw Hill
- ii) Beri (2008); *Business Statistics*; PVT Publishers New Delhi
- iii) Chandra J. S. (2003); *Statistical for Business and Economics*; Tata McGraw –Hall, New Delhi

Text Books for further Reading

- i) Mansfield Edwin; *Statistical for Business Economics Methods and Applications*; New York W.W Norton and Company
- ii) Enns Phillip G, (1985); *Business Statistical Methods ad Application*; Homewood Richard D Irwin Inc

Module compiled by Charles Karuga

Table of content

Course content.....	2
Table of content.....	3
CHAPTER 1: INTRODUCTION.....	4
CHAPTER 2: COLLECTION OF DATA.....	13
CHAPTER 3: ORGANIZATION AND REPRESENTATION OF DATA.....	27
CHAPTER 4: ANALYSIS AND INTERPRETATION OF DATA.....	47
CHAPTER 5: MEASURES OF DISPERSION.....	93
CHAPTER 6: CORRELATION AND REGRESSION ANALYSIS.....	122
CHAPTER 7: RANDOM VARIABLES AND THEIR PROBABILITY DISTRIBUTIONS.....	154
SAMPLE PAPERS.....	176
APPENDIX.....	190

LECTURE 1

CHAPTER 1: INTRODUCTION

Purpose

To introduce the student to the world of statistics and to acquaint them with the role of statistics in Business.



Objectives

- a) *Define statistics and explain its uses.*
- b) *Define Business Statistics*
- c) *State limitations of statistics.*
- d) *Explain why statistics is distrusted.*
- e) *Distinguish between descriptive and inferential statistics.*
- f) *Explain the types of variables.*
- g) *State the levels and scales of measurement.*

1.1 What is Statistics?

The word 'statistics' is defined by Croxton and Cowden as follows:-

"The collection, presentation, analysis and interpretation of the numerical data."

This definition clearly points out four stages in a statistical investigation, namely:

- 1) Collection of data 2) Presentation of data
- 3) Analysis of data 4) Interpretation of data

In addition to this, one more stage i.e. organization of data is suggested

Definition:

Business statistics is the science of good decision making in the face of uncertainty and is used in many disciplines such as financial analysis, econometrics, auditing, production and operations including services improvement, and marketing research..

1.2 Uses of Statistics

- a) To present the data in a concise and definite form : Statistics helps in classifying and tabulating raw data for processing and further tabulation for end users.
- b) To make it easy to understand complex and large data : This is done by presenting the data in the form of tables, graphs, diagrams etc., or by condensing the data with the help of means, dispersion etc.
- c) For comparison: Tables, measures of means and dispersion can help in comparing different sets of data..
- d) In forming policies: It helps in forming policies like a production schedule, based on the relevant sales figures. It is used in forecasting future demands.
- e) Enlarging individual experiences: Complex problems can be well understood by statistics, as the conclusions drawn by an individual are more definite and precise than mere statements on facts.
- f) In measuring the magnitude of a phenomenon:- Statistics has made it possible to count the population of a country, the industrial growth, the agricultural growth, the educational level (of course in numbers)

1.3 Limitations of Statistics

1. Statistics does not deal with individual measurements. Since statistics deals with aggregates of facts, it cannot be used to study the changes that have taken place in individual cases. For example, the wages earned by a single industry worker at any time, taken by itself is not a statistical datum. But the wages of workers of that industry can be used statistically. Similarly the marks obtained by Kamau of your class or the height of Atieno (also of your class) are not the subject matter of statistical study. But the average marks or the average height of your class has statistical relevance.

2. Statistics cannot be used to study qualitative phenomenon like morality, intelligence, beauty etc. as these cannot be quantified. However, it may be possible to analyze such problems statistically by expressing them numerically. For example we may study the intelligence of boys on the basis of the marks obtained by them in an examination.
3. Statistical results are true only on an average:- The conclusions obtained statistically are not universal truths. They are true only under certain conditions. This is because statistics as a science is less exact as compared to the natural science.
4. Statistical data, being approximations, are mathematically incorrect. Therefore, they can be used only if mathematical accuracy is not needed.
5. Statistics, being dependent on figures, can be manipulated and therefore can be used only when the authenticity of the figures has been proved beyond doubt..

1.3 Distrust of Statistics

It is often said by people that, "statistics can prove anything." There are three types of lies - lies, demand lies and statistics - wicked in the order of their naming. A Paris banker said, "Statistics is like a miniskirt, it covers up essentials but gives you the ideas."

Thus by "distrust of statistics" we mean lack of confidence in statistical statements and methods.

The following reasons account for such views about statistics.

1. Figures are convincing and, therefore people easily believe them.
2. They can be manipulated in such a manner as to establish foregone conclusions.
3. The wrong representation of even correct figures can mislead a reader. For example, John earned Ksh 400,000 in 1990 - 1991 and Jane earned Ksh 500,000. Reading this one would form the opinion that Jane is decidedly a better worker than John. However if we carefully examine the statement, we might reach a different conclusion as Jane's earning period is unknown to us. Thus while working with statistics one should not only avoid outright falsehoods but be alert to detect possible distortion of the truth.

1.4 Types of Statistics

Broadly speaking, statistics may be divided into two categories, ie descriptive and inferential statistics.

When analyzing data, for example, the marks achieved by 100 students for a piece of coursework, it is possible to use both descriptive and inferential statistics in your analysis of their marks. Typically, in most research conducted on groups of people, you will use both descriptive and inferential statistics to analyze your results and draw conclusions. *So what are descriptive and inferential statistics? And what are their differences?*

1.4.1 Descriptive Statistics

Descriptive statistics is the term given to the analysis of data that helps describe, show or summarize data in a meaningful way such that, for example, patterns might emerge from the data. Descriptive statistics do not, however, allow us to make conclusions beyond the data we have analyzed or reach conclusions regarding any hypotheses we might have made. They are simply a way to describe our data.

Descriptive statistics are very important, as if we simply presented our raw data it would be hard to visualize what the data was showing, especially if there was a lot of it. Descriptive statistics therefore allow us to present the data in a more meaningful way which allows simpler interpretation of the data. For example, if we had the results of 100 pieces of students' coursework, we may be interested in the overall performance of those students. We would also be interested in the distribution or spread of the marks. Descriptive statistics allow us to do this. Typically, there are two general types of statistic that are used to describe data:

- Measures of central tendency: these are ways of describing the central position of a frequency distribution for a group of data. In this case, the frequency distribution is simply the distribution and pattern of marks scored by the 100 students from the lowest to the highest. We can describe this central position using a number of statistics, including the mode, median, and mean.

- Measures of spread: these are ways of summarizing a group of data by describing how spread out the scores are. For example, the mean score of our 100 students may be 65 out of 100. However, not all students will have scored 65 marks. Rather, their scores will be spread out. Some will be lower and others higher. Measures of spread help us to summarize how spread out these scores are. To describe this spread, a number of statistics are available to us, including the range, quartiles, absolute deviation, variance and standard deviation.
- When we use descriptive statistics it is useful to summarize our group of data using a combination of tabulated description (i.e. tables), graphical description (i.e. graphs and charts) and statistical commentary (i.e. a discussion of the results).

1.4.2 Inferential Statistics

Whilst descriptive statistics examine our immediate group of data (for example, the 100 students' marks), inferential statistics aim to make inferences from this data in order to make conclusions that go beyond this data. In other words, inferential statistics are used to make inferences about a population from a sample in order to generalize (make assumptions about this wider population) and/or make predictions about the future.

For example, a Board of Examiners may want to compare the performance of 1000 students that completed an examination. Of these, 500 students are girls and 500 students are boys. The 1000 students represent our "population". Whilst we are interested in the performance of all 1000 students, girls and boys, it may be impractical to examine the marks of all of these students because of the time and cost required to collate all of their marks. Instead, we can choose to examine a "sample" of these students and then use the results to make generalizations about the performance of all 1000 students. For the purpose of our example, we may choose a sample size of 200 students. Since we are looking to compare boys and girls, we may randomly select 100 girls and 100 boys in our sample. We could then use this, for example, to see if there are any statistically significant differences in the mean mark between boys and girls, even though we have not measured all 1000 students.

1.5 Common Mistakes Committed In Interpretation of Statistics

1. Bias: - Bias means prejudice or preference of the investigator, which creeps in consciously and unconsciously in proving a particular point.
2. Generalization:- Some times on the basis of little data available one could jump to a conclusion, which leads to erroneous results.
3. Wrong conclusion:- The characteristics of a group if attached to an individual member of that group, may lead us to draw absurd conclusions.
4. Incomplete classification:- If we fail to give a complete classification, the influence of various factors may not be properly understood.
5. There may be a wrong use of percentages.
6. Technical mistakes may also occur.
7. An inconsistency in definition can even exist.
8. Wrong causal inferences may sometimes be drawn.

1.6 Types of Variables

1.6.1 Discrete Variable

A *discrete variable* is one that cannot take on all values within the limits of the variable. For example, responses to a five-point rating scale can only take on the values 1, 2, 3, 4, and 5. The variable cannot have the value 1.7. A variable such as a person's height can take on any value. Variables that can take on any value and therefore are not discrete are called continuous.

Statistics computed from discrete variables have many more possible values than the discrete variables themselves. The mean on a five-point scale could be 3.117 even though 3.117 is not possible for an individual score.

1.6.2 Continuous Variable

A continuous variable is one for which, within the limits the variable ranges, any value is possible. For example, the variable "Time to solve an anagram problem" is continuous since it could take 2 minutes, 2.13 minutes etc. to finish a problem. The variable "Number of correct

answers on a 100 point multiple-choice test" is not a continuous variable since it is not possible to get 54.12 problems correct. A variable that is not continuous is called "discrete"

1.7 Scales of measurement

1.7.1 Nominal Scale

Nominal measurement consists of assigning items to groups or categories. No quantitative information is conveyed and no ordering of the items is implied. Nominal scales are therefore qualitative rather than quantitative. Religious preference, race, and sex are all examples of nominal scales. Frequency distributions are usually used to analyze data measured on a nominal scale. The main statistic computed is the mode. Variables measured on a nominal scale are often referred to as categorical or qualitative variables.

1.7.2 Ordinal Scale

Measurements with ordinal scales are ordered in the sense that higher numbers represent higher values. However, the intervals between the numbers are not necessarily equal. For example, on a five-point rating scale measuring attitudes toward gun control, the difference between a rating of 2 and a rating of 3 may not represent the same difference as the difference between a rating of 4 and a rating of 5. There is no "true" zero point for ordinal scales since the zero point is chosen arbitrarily. The lowest point on the rating scale in the example was arbitrarily chosen to be 1. It could just as well have been 0 or -5.

1.7.3 Interval Scale

On interval measurement scales, one unit on the scale represents the same magnitude on the trait or characteristic being measured across the whole range of the scale. For example, if anxiety were measured on an interval scale, then a difference between a score of 10 and a score of 11 would represent the same difference in anxiety as would a difference between a score of 50 and a score of 51. Interval scales do not have a "true" zero point, however, and therefore it is not possible to make statements about how many times higher one score is than another. For the anxiety scale, it would not be valid to say that a person with a score of 30 was twice as anxious

as a person with a score of 15. True interval measurement is somewhere between rare and nonexistent in the behavioral sciences. No interval-level scale of anxiety such as the one described in the example actually exists. A good example of an interval scale is the Fahrenheit scale for temperature. Equal differences on this scale represent equal differences in temperature, but a temperature of 30 degrees is not twice as warm as one of 15 degrees.

1.7.4 Ratio Scale

Ratio scales are like interval scales except they have true zero points. A good example is the Kelvin scale of temperature. This scale has an absolute zero. Thus, a temperature of 300 Kelvin is twice as high as a temperature of 150 Kelvin.



Chapter Review Questions

- 1. Define statistics***
- 2. Define Business statistics***
- 3. Explain how the knowledge of statistics may be applied in business situation.***
- 4. State two ways in which statistics may be misused.***
- 5. Distinguish between descriptive and inferential statistics.***
- 6. State the scale of measurement the following can be classified into***
 - i. The mass of a bull***
 - ii. The length of time spent in a restaurant***
 - iii. The rank of an army officer***
 - iv. The type of vehicle driven by a celebrity.***

Reference

- i. *Research methods* by Mugenda Olive M and Mugenda Abel G. Pg13-14, Pg 71-93
- ii. *Business Calculations and statistics simplified by* N.A Saleemi. Revised Edition. Pg 237-244
- iii. *Essentials of statistics for Business and Economics* by Anderson Sweety Williams Pg 5-13

LECTURE 2

CHAPTER 2: COLLECTION OF DATA



Purpose

To introduce the student to some basic techniques of collection of data.

Objectives

- a) Distinguish between primary and secondary data.*
- b) Describe different methods of data collection.*
- c) Define sampling and explain various methods of sampling*

For any statistical enquiry, the basic objective is to collect facts and figures relating to a particular phenomenon for further statistical analysis. The process of counting, enumeration or measurement together with systematic recording of results is called collection of statistical data.

2.1 Primary and Secondary Data

Primary data is data that you collect yourself using such methods as:

- **direct observation** - lets you focus on details of importance to you; lets you see a system in *real* rather than *theoretical* use (other faults are unlikely or trivial in theory but quite real and annoying in practice);
- **surveys** - written surveys let you collect considerable quantities of detailed data. You have to either trust the honesty of the people surveyed or build in self-verifying questions (e.g. questions 9 and 24 ask basically the same thing but using different words - different answers may indicate the surveyed person is being inconsistent, dishonest or inattentive).
- **interviews** - slow, expensive, and they take people away from their regular jobs, but they allow in-depth questioning and follow-up questions. They also show non-verbal communication such as face-pulling, fidgeting, shrugging, hand gestures, sarcastic expressions that add further meaning to spoken words. e.g. "I think it's a GREAT system"

could mean vastly different things depending on whether the person was sneering at the time! A problem with interviews is that people might say what they think the interviewer wants to hear; they might avoid being honestly critical in case their jobs or reputation might suffer.

- **logs** (e.g. fault logs, error logs, complaint logs, transaction logs). Good, empirical, objective data sources (usually, if they are used well). Can yield lots of valuable data about system performance over time under different conditions.

Primary data can be relied on because you know where it came from and what was done to it. It's like cooking something yourself. You know what went into it.

Secondary data is collected from external sources such as:

- TV, radio, internet
- magazines, newspapers
- reviews
- research articles
- stories told by people you know

There's a lot more secondary data than primary data, and secondary data is a *whole lot* cheaper and easier to acquire than primary data. The problem is that often the reliability, accuracy and integrity of the data is uncertain. Who collected it? Can they be trusted? Did they do any preprocessing of the data? Is it biased? How old is it? Where was it collected? Can the data be verified, or does it have to be taken on faith?

Often secondary data has been pre-processed to give totals or averages and the original details are lost so you can't verify it by replicating the methods used by the original data collectors.

In short, primary data is expensive and difficult to acquire, but it's trustworthy. Secondary data is cheap and easy to collect, but must be treated with caution.

2.2 Methods of collecting Primary data

In primary data collection, you collect the data yourself using methods such as interviews and questionnaires. The key point here is that the data you collect is unique to you and your research and, until you publish, no one else has access to it.

There are many methods of collecting primary data and the main methods include:

- questionnaires
- interviews
- focus group interviews
- observation
- case-studies
- diaries
- critical incidents
- Portfolios.

The primary data, which is generated by the above methods, may be qualitative in nature (usually in the form of words) or quantitative (usually in the form of numbers or where you can make counts of words used). We briefly outline these methods but you should also read around the various methods. A list of suggested research methodology texts is given in your *Module Study Guide* but many texts on social or educational research may also be useful and you can find them in your libra

2.2.1 Questionnaires

Questionnaires are a popular means of collecting data, but are difficult to design and often require many rewrites before an acceptable questionnaire is produced.

Advantages:

- Can be used as a method in its own right or as a basis for interviewing or a telephone survey.

- Can be posted, e-mailed or faxed.
- Can cover a large number of people or organisations.
- Wide geographic coverage.
- Relatively cheap.
- No prior arrangements are needed.
- Avoids embarrassment on the part of the respondent.
- Respondent can consider responses.
- Possible anonymity of respondent.
- No interviewer bias.

Disadvantages:

- Design problems.
- Questions have to be relatively simple.
- Historically low response rate (although inducements may help).
- Time delay whilst waiting for responses to be returned.
- Require a return deadline.
- Several reminders may be required.
- Assumes no literacy problems.
- No control over who completes it.
- Not possible to give assistance if required.

- Problems with incomplete questionnaires.
- Replies not spontaneous and independent of each other.
- Respondent can read all questions beforehand and then decide whether to complete or not. For example, perhaps because it is too long, too complex, uninteresting, or too personal

2.2.2 Interviews

Interviewing is a technique that is primarily used to gain an understanding of the underlying reasons and motivations for people's attitudes, preferences or behaviour. Interviews can be undertaken on a personal one-to-one basis or in a group. They can be conducted at work, at home, in the street or in a shopping centre, or some other agreed location.

Personal interview

Advantages:

- Serious approach by respondent resulting in accurate information.
- Good response rate.
- Completed and immediate.
- Possible in-depth questions.
- Interviewer in control and can give help if there is a problem.
- Can investigate motives and feelings.
- Can use recording equipment.
- Characteristics of respondent assessed – tone of voice, facial expression, hesitation, etc.
- Can use props.

- If one interviewer used, uniformity of approach.
- Used to pilot other methods.

Disadvantages:

- Need to set up interviews.
- Time consuming.
- Geographic limitations.
- Can be expensive.
- Normally need a set of questions.
- Respondent bias – tendency to please or impress, create false personal image, or end interview quickly.
- Embarrassment possible if personal questions.
- Transcription and analysis can present problems – subjectivity.
- If many interviewers, training required.

2.2.3 Case-studies

The term case-study usually refers to a fairly intensive examination of a single unit such as a person, a small group of people, or a single company. Case-studies involve measuring what is there and how it got there. In this sense, it is historical. It can enable the researcher to explore, unravel and understand problems, issues and relationships. It cannot, however, allow the researcher to generalize, that is, to argue that from one case-study the results, findings or theory developed apply to other similar case-studies. The case looked at may be unique and, therefore not representative of other instances. It is, of course, possible to look at several case-studies to represent certain features of management that we are interested in studying. The case-study

approach is often done to make practical improvements. Contributions to general knowledge are incidental.

The case-study method has four steps:

1. Determine the present situation.
2. Gather background information about the past and key variables.
3. Test hypotheses. The background information collected will have been analysed for possible hypotheses. In this step, specific evidence about each hypothesis can be gathered. This step aims to eliminate possibilities which conflict with the evidence collected and to gain confidence for the important hypotheses. The culmination of this step might be the development of an experimental design to test out more rigorously the hypotheses developed, or it might be to take action to remedy the problem.
4. Take remedial action. The aim is to check that the hypotheses tested actually work out in practice. Some action, correction or improvement is made and a re-check carried out on the situation to see what effect the change has brought about.

The case-study enables rich information to be gathered from which potentially useful hypotheses can be generated. It can be a time-consuming process. It is also inefficient in researching situations which are already well structured and where the important variables have been identified. They lack utility when attempting to reach rigorous conclusions or determining precise relationships between variables.

2.2.4 Diaries

A diary is a way of gathering information about the way individuals spend their time on professional activities. They are not about records of engagements or personal journals of thought! Diaries can record either quantitative or qualitative data, and in management research can provide information about work patterns and activities.

Advantages:

- Useful for collecting information from employees.
- Different writers compared and contrasted simultaneously.
- Allows the researcher freedom to move from one organization to another.
- Researcher not personally involved.
- Diaries can be used as a preliminary or basis for intensive interviewing.
- Used as an alternative to direct observation or where resources are limited.

Disadvantages:

- Subjects need to be clear about what they are being asked to do, why and what you plan to do with the data.
- Diarists need to be of a certain educational level.
- Some structure is necessary to give the diarist focus, for example, a list of headings.
- Encouragement and reassurance are needed as completing a diary is time-consuming and can be irritating after a while.
- Progress needs checking from time-to-time.
- Confidentiality is required as content may be critical.
- Analyses problems, so you need to consider how responses will be coded before the subjects start filling in diaries.

2.2.5 Portfolios

A measure of a manager's ability may be expressed in terms of the number and duration of 'issues' or problems being tackled at any one time. The compilation of problem portfolios is

recording information about how each problem arose, methods used to solve it, difficulties encountered, etc. This analysis also raises questions about the person's use of time. What proportion of time is occupied in checking; in handling problems given by others; on self-generated problems; on 'top-priority' problems; on minor issues, etc? The main problem with this method and the use of diaries is getting people to agree to record everything in sufficient detail for you to analyze. It is very time-consuming!

2.3 Sampling

Collecting data is time consuming and expensive, even for relatively small amounts of data. Hence, it is highly unlikely that a complete population will be investigated. Because of the time and cost elements the amount of data you collect will be limited and the number of people or organizations you contact will be small in number. You will, therefore, have to take a sample and usually a small sample.

Sampling theory says a correctly taken sample of an appropriate size will yield results that can be applied to the population as a whole. There is a lot in this statement but the two fundamental questions to ensure generalization are:

1. How is a sample taken correctly?
2. How big should the sample be?

The answer to the second question is 'as large as possible given the circumstances'. It is like answering the question 'How long is a piece of string'? It all depends on the circumstances.

Whilst we do not expect you to normally generalize your results and take a large sample, we do expect that you follow a recognized sampling procedure, such that, if the sample was increased generalization would be possible. You therefore need to know some of the basics of sampling. This will be done by reference to the following example.

The theory of sampling is based on random samples – where all items in the population have the same chance of being selected as sample units. Random samples can be drawn in a number of

ways but are usually based on having some information about population members. This information is usually in the form of an alphabetical list – called the sampling frame.

Three types of random sample can be drawn – a simple random sample (SRS), a stratified sample and a systematic sample.

2.3.1 Simple random sampling

Simple random sampling can be carried out in two ways – the lottery method and using random numbers.

The **lottery method** involves:

- transferring each person's name from the list and putting it on a piece of paper
- the pieces of paper are placed in a container and thoroughly mixed
- the required number are selected by someone without looking
- the names selected are the simple random sample.

This is basically similar to a game of bingo or the national lottery. This procedure is easy to carry out especially if both population and sample are small, but can be tedious and time consuming for large populations or large samples.

Alternatively **random numbers** can be used. Random numbers are strings of digits that have been generated by the lottery method and can be found in books of statistical tables. An example of these is:

03	47	43	73	86	36	96	47	36	61
97	74	24	67	62	42	81	14	57	20
16	76	62	27	66	56	50	26	71	07
12	56	85	99	26	96	96	68	27	31

55 59 56 35 64 38 04 80 46 22

Random numbers tend to be written in pairs and blocks of 5 by 5 to make reading easy. However, care is needed when reading these tables. The numbers can be read in any direction but they should be read as a single string of digits i.e. left to right as 0, 3, 4, 7 etc', or top to bottom as 0, 9, 1, 1, 5, 3, 7, ... etc. It is usual to read left to right.

The **random number** method involves:

- Allocating a number to each person on the list (each number must consist of the same number of digits so that the tables can be read consistently).
- Find a starting point at random in the tables (close your eyes and point).
- Read off the digits.
- The names matching the numbers are the sample units.

For the example of selecting nine people at random from 90:

- a) The sampling frame is the list of 90 people. Number this list 00, 01, 02, ..., 89. Note that each number has two digits and the numbering starts from 00.
- b) Suppose a starting point is found at random from the random number tables and let this number be 16. Then the person that has been numbered 16 is the first sample unit.
- c) Let the next two digits be 76, then the person numbered 76 is the second sample unit.

This procedure is repeated until the nine people have been identified.

- d) Any number occurring for second time is ignored as is any two-digit number over 89.

Simple random number sampling is used as the basis for many other sampling methods, but has two disadvantages:

- A sampling frame is required. This may not be available, exist or be incomplete.
- The procedure is unbiased but the sample may be biased. For instance, if the 90 people are a mixture of men and women and all men were selected this would be a biased sample.

2.3.2 Stratified Sampling

To overcome the second problem above, a stratified sample can be taken. In this the population structure is reflected in the sample structure, with respect to some criterion.

For example, suppose the 90 people consist of 30 men and 60 women. If gender is the criterion for stratification then:

$\frac{30}{90}$ of the sample should be men

i.e. $\frac{30}{90} \times 9 = 3men$

$\frac{60}{90}$ of the sample should be women

i.e. $\frac{60}{90} \times 9 = 3 \text{ women}$

Thus the sample reflects the population structure in terms of gender.

The three men and six women would then be selected by simple random sampling e.g., random numbers.

The problem with this approach is the criterion for stratification, (e.g., age, sex, job description), is chosen by you – it is subjective and may not be the best or more appropriate criterion. Also a more detailed sampling frame is required.

2.3.3 Systematic sampling

Whilst not truly random this is a method that is used extensively because it is easy to operate and quick, even when the population and the sample are large. For example, for the population 90 and sample of nine:

Split the sampling frame in to nine equal groups.

i.e. 1 to 9

10 to 19

Etc

80 to 89

Select a number between 1 and 9 using random number tables.

Suppose this number is 6.

Person numbered 6 is chosen.

Then the 16th, 26th, 36th, 46th, 56th, 66th, 76th, and 86th people are the remaining sample units.

If no sampling frame is available access to the population is necessary, such as customers of a business such as a leisure centre, restaurant or museum.

Systematic sampling can be used by selecting a random number say 25.

Then the 25th person to enter is the first sample unit.

The 50th person to enter is the second sample unit.

This process is carried on until the required sample size is met.

This approach usually generates a good cross section of the population. However, you may need a team of people when no sampling frame exists to help with counting, interviewing, etc.



Review Questions

1. *Distinguish between primary and secondary data.*
2. *Discuss the various methods of data collection. Indicate the situations in which each of these methods should be used.*
3. *What is sampling?*
4. *State four reasons why is it important to study a sample instead of the whole population*
5. *Discuss the various sampling methods.*

- a) *Research methods* by Mugenda Olive M and Mugenda Abel G. Pg 71-93
- b) *Business Calculations and statistics simplified* by N.A Saleemi. Revised Edition. Pg 246-25

LECTURE 3

CHAPTER 3: ORGANIZATION AND REPRESENTATION OF DATA



Purpose

To introduce the student to various techniques of organizing and presenting sets of data.

Objectives

- a) *Explain the general principle of constructing diagrams.*
- b) *Represent information in form of*
 - i. *Bar charts*
 - ii. *Histograms*
 - iii. *Pie charts*
 - iv. *Frequency polygons*
 - v. *Ogives*
 - vi. *stem and leaf diagrams*
 - vii. *Box and whisker plots*

3.1 Introduction

Graphs and diagram leave a lasting impression on the mind and make intelligible and easily understandable the salient features of the data. Forecasting also becomes easier with the help of graph. Thus it is of interest to study the graphical representation of data.

3.2 General Principles of Constructing Diagrams

1. The diagrams should be simple.
2. Each diagram must be given a clear, concise and suitable title without damaging clarity.
3. A proper proportion between height and width must be maintained in order to avoid an unpleasant look.

4. Select a proper scale; it should be in even numbers or in multiples of five or ten.

e.g. 25, 50, 75 or 10, 20, 30, 40... etc. But no fixed rule.

5. In order to clear certain points, always put footnotes.
 6. An index, explaining different lines, shades and colors should be given.
 7. Diagrams should be absolutely neat and clean.
- "The important point that must be borne in mind at all times that the pictorial representation chosen for any situation must depict the true relationship and point out the proper conclusion. Above all the chart must be honest.".... C. W. LOWE.

3.3 Bar Diagrams

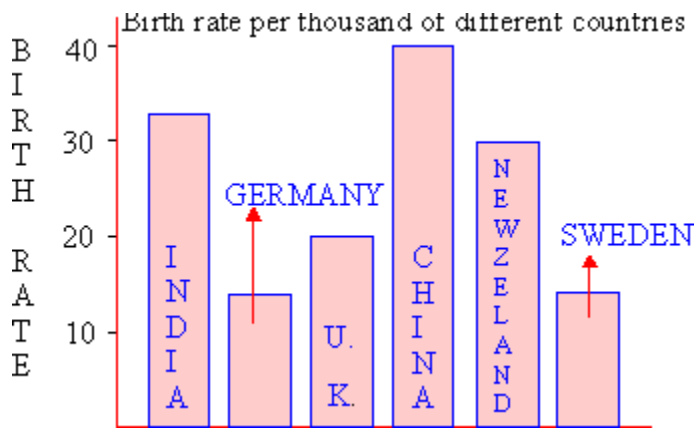
3.3.1 Simple 'Bar diagram'

It represents only one variable. For example sales, production, population figures etc. for various years may be shown by simple bar charts. Since these are of the same width and vary only in heights (or lengths), it becomes very easy for readers to study the relationship. Simple bar diagrams are very popular in practice. A bar chart can be either vertical or horizontal; vertical bars are more popular.

Illustration :- The following table gives the birth rate per thousand of different countries over a certain period of time.

Country	Birth rate	Country	Birth rate
India	33	China	40
Germany	15	New Zealand	30
U. K.	20	Sweden	15

Represent the above data by a suitable diagram.



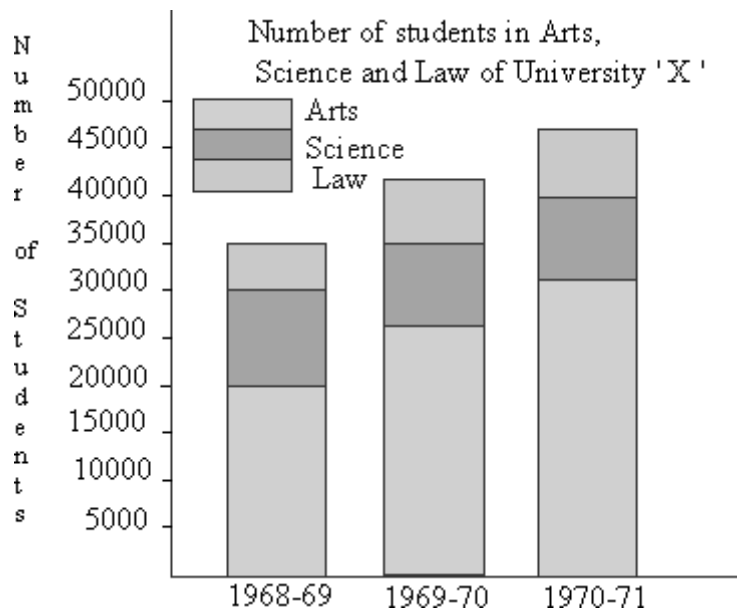
Comparing the size of bars, you can easily see that China's birth rate is the highest while Germany and Sweden equal in the lowest positions. Such diagrams are also known as component bar diagrams.

3.3.2 Sub - divided Bar Diagram

While constructing such a diagram, the various components in each bar should be kept in the same order. A common and helpful arrangement is that of presenting each bar in the order of magnitude with the largest component at the bottom and the smallest at the top. The components are shown with different shades or colors with a proper index.

Illustration:- During 1968 - 71, the number of students in University ' X ' are as follows. Represent the data by a similar diagram.

Year	Arts	Science	Law	Total
1968-69	20,000	10,000	5,000	35,000
1969-70	26,000	9,000	7,000	42,000
1970-71	31,000	9,500	7,500	48,000



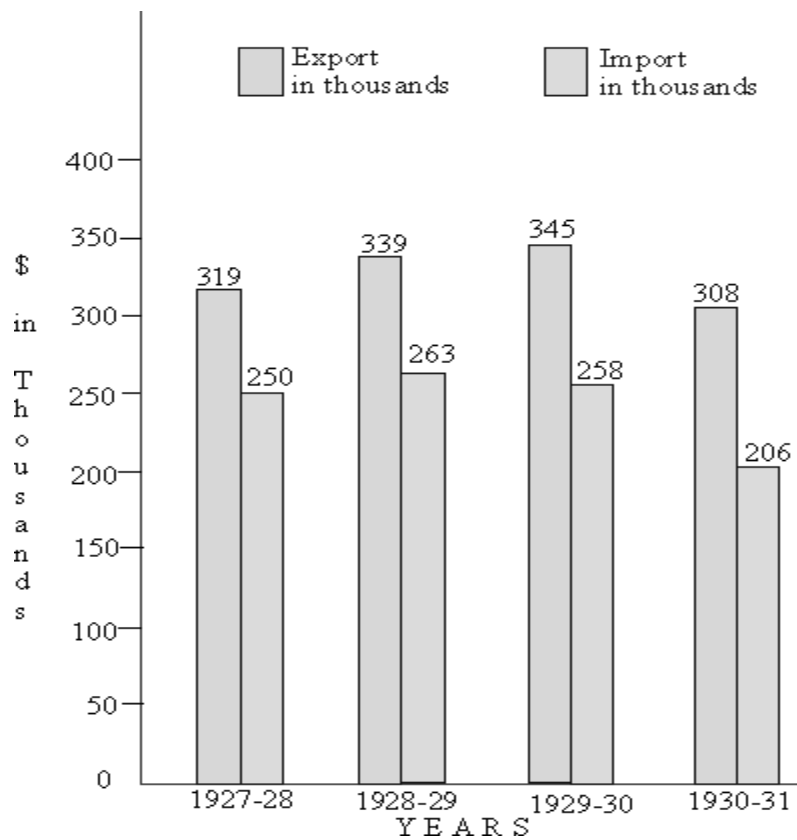
3.3.3 Multiple Bar Diagram

This method can be used for data which is made up of two or more components. In this method the components are shown as separate adjoining bars. The height of each bar represents the actual value of the component. The components are shown by different shades or colors. Where changes in actual values of component figures only are required, multiple bar charts are used.

Illustration:- The table below gives data relating to the exports and imports of a certain country X (in thousands of dollars) during the four years ending in 1930 - 31.

Year	Export	Import
1927 - 28	319	250
1928 - 29	339	263
1929 - 30	345	258
1930 - 31	308	206

Represent the data by a suitable diagram



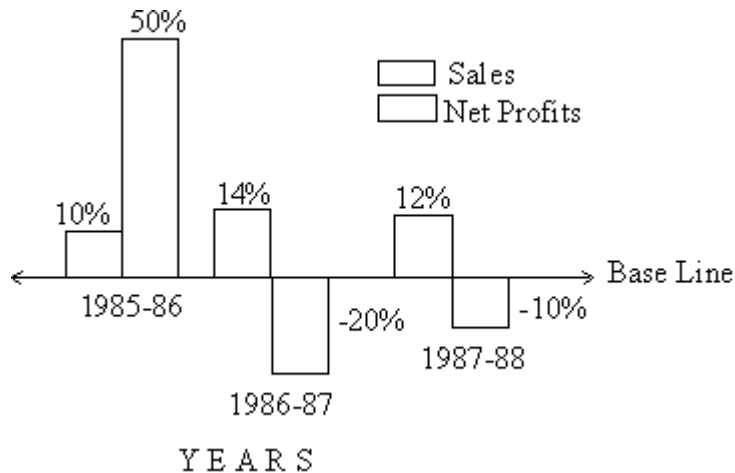
3.3.4 Deviation Bar Charts

Deviation bars are used to represent net quantities - excess or deficit i.e. net profit, net loss, net exports or imports, swings in voting etc. Such bars have both positive and negative values. Positive values lie above the base line and negative values lie below it.

Illustration:-

Years	Sales	Net profits
1985 - 86	10%	50%
1986 - 87	14%	-20
1987 - 88	12%	-10%

Present the above data by a suitable diagram showing the sales and net profits of private industrial companies.



3.4 Pie Chart

i) Geometrically it can be seen that the area of a sector of a circle taken radially, is proportional to the angle at its center. It is therefore sufficient to draw angles at the center, proportional to the original figures. This will make the areas of the sector proportional to the basic figures.

For example, let the total be 1000 and one of the component be 200, then the angle will be

$$\left(\frac{200}{1000} \right) \times 360^\circ = 72^\circ$$

In general, angle of sector at the center corresponding to a component

$$= \left(\frac{\text{Component}}{\text{Total}} \right) \times 360^\circ$$

ii) When a statistical phenomenon is composed of different components which are numerous (say four or more components), bar charts are not suitable to represent them because, under this situation, they become very complex and their visual impressions are questioned. A pie diagram is suitable for such situations. It is a circular diagram which is a circle (pie) divided by the radii, into sectors (like slices of a cake or pie). The area of a sector is proportional to the size of each

Pie charts are useful to compare different parts of a whole amount. They are often used to present financial information. E.g. A Company's expenditure can be shown to be the sum of its parts including different expense categories such as salaries, borrowing interest, taxation and general running costs (i.e. rent, electricity, heating etc).

A pie chart is a circular chart in which the circle is divided into sectors. Each sector visually represents an item in a data set to match the amount of the item as a percentage or fraction of the total data set.

Example 9

A family's weekly expenditure on its house mortgage, food and fuel is as follows:

Expense	\$
Mortgage	300
Food	225
Fuel	75

Draw a pie chart to display the information.

Solution:

$$\begin{aligned} \text{The total weekly expenditure} &= \$300 + \$225 + \$75 \\ &= \$600 \end{aligned}$$

We can find what percentage of the total expenditure each item equals.

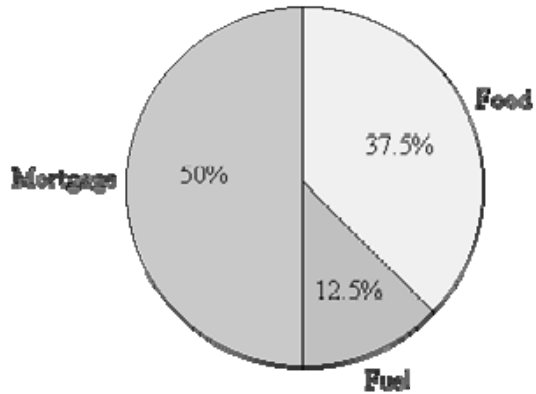
Percentage of weekly expenditure on:

$$\text{Mortgage} = \frac{300}{600} \times 100\% = 50\%$$

$$\text{Food} = \frac{225}{600} \times 100\% = 37.5\%$$

$$\text{Fuel} = \frac{75}{600} \times 100\% = 12.5\%$$

To draw a pie chart, divide the circle into 100 percentage parts. Then allocate the number of percentage parts required for each item.



- It is simple to read a pie chart. Just look at the required sector representing an item (or category) and read off the value. For example, the weekly expenditure of the family on food is 37.5% of the total expenditure measured.
- A pie chart is used to compare the different parts that make up a whole amount.

3.5 Graphs

A graph is a visual representation of data by a continuous curve on a squared (graph) paper. Like diagrams, graphs are also attractive, and eye-catching, giving a bird's eye-view of data and revealing their inner pattern.

Graphs of Frequency Distributions:-

The methods used to represent a grouped data are :-

1. Histogram
2. Frequency Polygon
3. Frequency Curve
4. Ogive or Cumulative Frequency Curve

3.5.1 Histogram

It is defined as a pictorial representation of a grouped frequency distribution by means of adjacent rectangles, whose areas are proportional to the frequencies.

To construct a Histogram, the class intervals are plotted along the x-axis and corresponding frequencies are plotted along the y - axis. The rectangles are constructed such that the height of each rectangle is proportional to the frequency of the class and width is equal to the length of the class. If all the classes have equal width, then all the rectangles stand on the equal width. In case of classes having unequal widths, rectangles too stand on unequal widths (bases). For open-classes, Histogram is constructed after making certain assumptions. As the rectangles are adjacent leaving no gaps, the class-intervals become of the inclusive type, adjustment is necessary for end points only.

For example, in a book sale, you want to determine which books were most popular, the high priced books, the low priced books, books most neglected etc. Let us say you sold total 31 books at this book-fair at the following prices.

\$2, \$ 1, \$ 2, \$ 2, \$ 3, \$ 5, \$ 6, \$ 17, \$ 17, \$ 7, \$ 15, \$ 7, \$ 7, \$ 18, \$ 8, \$ 10, \$ 10, \$ 9, \$ 13, \$ 11, \$ 12, \$ 12, \$ 12, \$ 14, \$ 16, \$ 18, \$ 20, \$ 24, \$ 21, \$ 22, \$ 25.

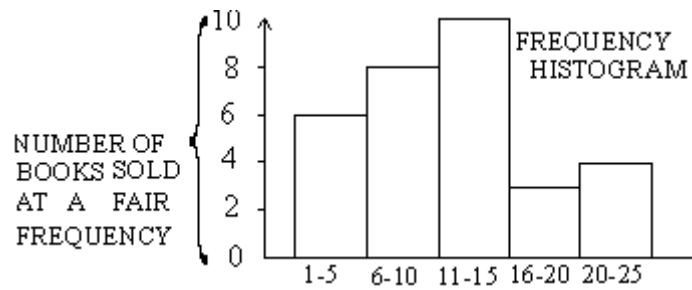
The books are ranging from \$1 to \$25. Divide this range into number of groups, class intervals. Typically, there should not be fewer than 5 and more than 20 class-intervals are best for a frequency Histogram.

Our first class-interval includes the lowest price of the data and, the last-interval of course includes, the highest price. Also make sure that overlapping is avoided, so that, no one price falls into two class-intervals. For example you have class intervals as 0-5, 5-10, 10-15 and so on, then the price \$10 falls in both 5-10 and 10-15. Instead if we use \$1 - \$5, \$6=\$10, the class-intervals will be mutually exclusive.

Therefore now we have distribution of books at a book-faire

Class-interval	Frequency
\$ 1- \$ 5	6
\$6 - \$10	8
\$11 - \$15	10
\$16 - \$20	3
\$21 - \$25	4
Total	$n = \sum f_i = 31$

Note that each class-interval is of equal width i.e. \$5 inclusive. Now we draw the frequency Histogram as under.

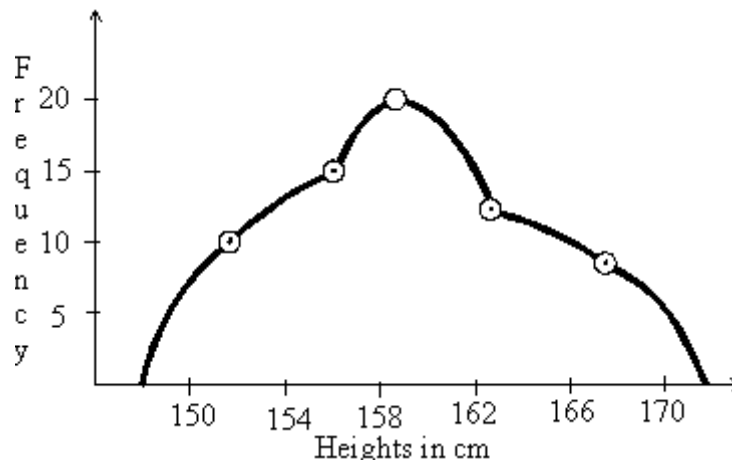


LECTURE 5

Methods of Presenting Data (continuation from lecture 4)

3.5.2 Frequency Distribution (Curve):-

Frequency distribution curves are like frequency polygons. In frequency distribution, instead of using straight line segments, a smooth curve is used to connect the points. The frequency curve for the above data is shown as:



3.5.3 Ogives or Cumulative Frequency Curves

When frequencies are added, they are called cumulative frequencies. The curve obtained by plotting cumulating frequencies is called a cumulative frequency curve or an ogive (pronounced ojive).

To construct an Ogive:-

- 1) Add up the progressive totals of frequencies, class by class, to get the cumulative frequencies.
- 2) Plot classes on the horizontal (x-axis) and cumulative frequencies on the vertical (y-axis).

3) Join the points by a smooth curve. Note that Ogives start at (i) zero on the vertical axis, and (ii) outside class limit of the last class. In most of the cases it looks like 'S'.

Note that cumulative frequencies are plotted against the 'limits' of the classes to which they refer.

(A) Less than Ogive:- To plot a less than ogive, the data is arranged in ascending order of magnitude and the frequencies are cumulated starting from the top. It starts from zero on the y-axis and the lower limit of the lowest class interval on the x-axis.

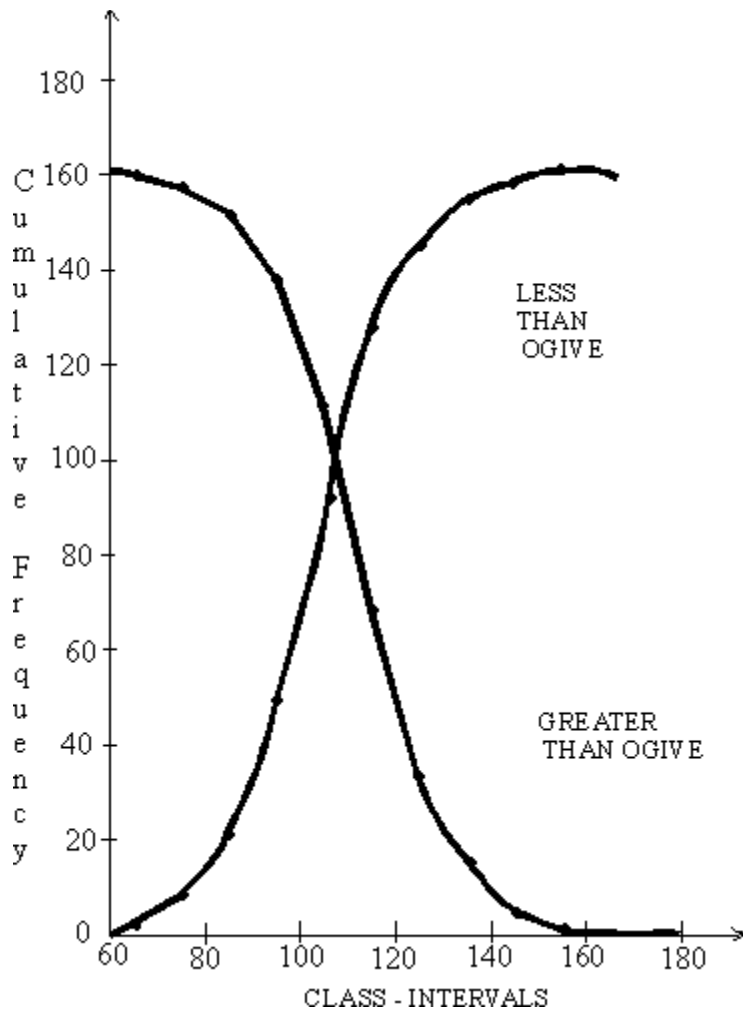
(B) Greater than Ogive:- To plot this ogive, the data are arranged in the ascending order of magnitude and frequencies are cumulated from the bottom. This curve ends at zero on the the y-axis and the upper limit of the highest class interval on the x-axis.

Illustrations:- On a graph paper, draw the two ogives for the data given below of the I.Q. of 160 students.

Class -intervals:	60 - 70	70 - 80	80 - 90	90 - 100	100 - 110
No. of students:	2	7	12	28	42

110 - 120	120 - 130	130 - 140	140 - 150	150 - 160
36	18	10	4	1

Solution classes	f	c. f. less than	c. f. greater than
60 - 70	2	2	159+1 = 160
70 - 80	7	2+7 = 9	151+7 = 158
80 - 90	12	9+12 = 21	139+12 = 151
90 -100	28	21+28 = 49	111+28 = 139
100 -110	42	49+42 = 91	69+42 = 111
110 -120	36	91+36 = 127	33+36 = 69
120 -130	18	127+18 = 145	15+18 = 33
130 -140	10	145+10 = 155	5+10 = 15
140 -150	4	155+4 = 159	1+4 = 5
150 -160	1	159+1 = 160	1
	$\Sigma f = 160$		



Uses: - Certain values like median, quartiles, deciles, quartile deviation, coefficient of skewness etc. can be located using ogives. It can be used to find the percentage of items having values less than.

3.6 Stem and Leaf Diagram

A stem and leaf diagram provides a **visual summary** of your data. This diagram provides a **partial sorting** of the data and allows you to detect the **distributional pattern** of the data.

There are three steps for drawing a stem and leaf diagram.

1. Split the data into two pieces, stem and leaf.
2. Arrange the stems from low to high.
3. Attach each leaf to the appropriate stem.

Example

Suppose you have the heights of 20 people as follows:

154, 143, 148, 139, 143, 147, 153, 162, 136, 147, 144, 143, 139, 142, 143, 156, 151, 164, 157, 149, 146

What we have here is almost a stem and leaf diagram. Note that with the data written in this way you can see what the modal class is (the one with the most values. You can also see the shape of the distribution- most of the values are in the 140s with higher or lower values rarer.

To change this into a stem and leaf diagram, we just simplify it a little. Instead of writing out the full figures each time (143, 143, 144, 143, ...) we write '14' and call this the 'stem' and then write 3, 3, 4, 3, ... (these being the 'leaves'). We would usually, however, write the leaves in order (with the smallest first). Finally, we must also include a little key so that people know how to interpret the diagram.

So we finish up with:

Stem	Leaf
13	6, 9, 9
14	2, 3, 3, 3, 3, 4
14	6, 7, 7, 8, 9
15	1, 3, 4
15	6, 7
16	2, 4

Key:

13 | 6 means 136

3.6.1 Back-to-back stem and leaf diagram

Back-to-back stem plots are used to compare two distributions side-by-side. This type of double stem plot contains three columns, each separated by a vertical line. The center column contains the stems. The first and third columns each contain the leaves of a different distribution. The numbers for the leaves of the distribution in the leftmost column are aligned to the right and are listed in increasing order from right to left. Here is an example of a back-to-back stem plot comparing the distribution of marks obtained in an exam by a sample of 25 boys and 25 girls.

BOYS						GIRLS						
				3	4	40	5	4	1	2	8	5
			3	5	5	50	2	3	5	8	9	4
2	2	3	3	4	5	60	3	5	6	4	5	
5	5	2	8	0	2	70	0	3	3			
			3	1	3	80	3	6	4			
				4	4	90	3	4				

KEY: 40 | 5 =45

Can you comment on the shape of the distribution of the two sets of data?

3.7 Box and Whisker Plots

It is one step further to stem-and-leaf. It displays a number of statistics like, median, lower quartile (Q1), upper quartile (Q3), Inter-quartile range (IQR). It tells us about the symmetry of the distribution and also gives us the idea about the highest and the lowest values.

Example

Statistics CAT scores of 12 students are as follows:-

10, 22, 24, 27, 31, 33, 39, 40, 42, 43, 44, 45

Draw a box and whisker plot to represent the above scores.

Solution: The scores are arranged in the ascending order. 10, 22, 24, 27, 31, 33, 39, 40, 42, 43, 44, 45

1) Since $n = 12$ (total items)

the two middle scores are $\frac{12}{2} = 6^{th}$ and $\frac{12+2}{2} = 7^{th}$

i.e. 33 and 39 respectively.

Therefore the average of the two is the median

i.e. Median = $\frac{1}{2}[33 + 39] = \frac{1}{2}[72] = 36$

2) The quartile (Q1) is the median of the bottom half. i.e. 25th percentile.

Thus

$$Q_1 = \frac{n}{4} = \frac{12}{4} = 3^{\text{rd}} \text{ score} = 24$$

- 3) The upper quartile (Q_3) is the median of the top half. i.e. 75th percentile.

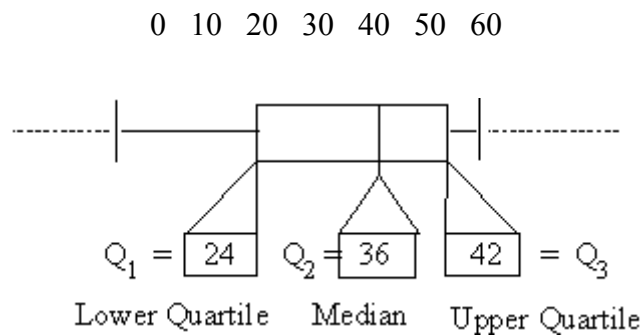
Thus

$$Q_3 = \frac{3N}{4} = \frac{3(12)}{4} = 9^{\text{th}} \text{ score} = 42$$

Now the box-plot is constructed as follows:-

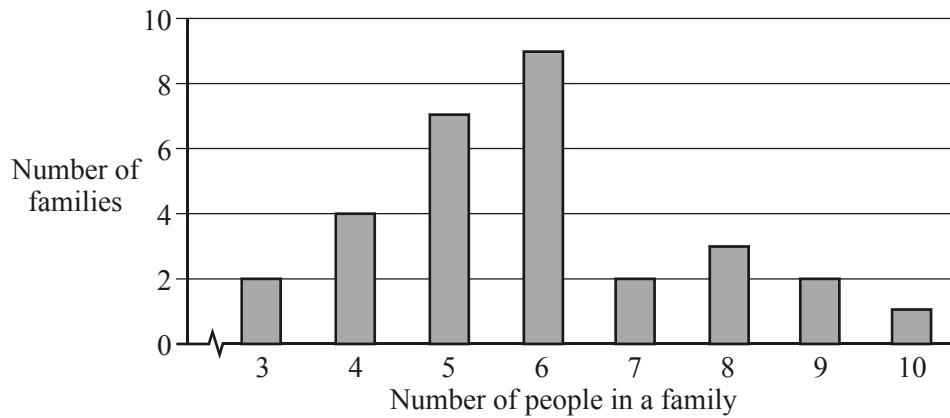
- i) The line inside the box indicates the median.
- ii) The left side of this box indicates the lower quartile (Q_1).
- iii) The right side of this box indicates the upper quartile (Q_3).
- iv) A straight line is then drawn from the lowest value of this distribution through the box to the highest value of this distribution. This horizontal straight line is called the "Whiskers".

Then the above CAT score in box-plot will look like this:



Chapter Review Questions

1. The bar chart below shows the number of people in a selection of families.



(a) How many families are represented?

(b) Write down the mode of the distribution.

(c) Find, correct to the nearest whole number, the mean number of people in a family.

2. A marine biologist records as a frequency distribution the lengths (L), measured to the nearest centimetre, of 100 mackerel. The results are given in the table below.

<i>Length of mackerel (L cm)</i>	<i>Number of mackerel</i>
$27 < L \leq 29$	2
$29 < L \leq 31$	4
$31 < L \leq 33$	8
$33 < L \leq 35$	21
$35 < L \leq 37$	30
$37 < L \leq 39$	18

$39 < L \leq 41$	12
$41 < L \leq 43$	5
	100

(a) Construct a cumulative frequency table for the data in the table.

(b) Draw a cumulative frequency curve.

Hint: Plot your cumulative frequencies at the top of each interval.

3. The following table shows the age distribution of teachers who smoke at Fegi High School.

<i>Ages</i>	<i>Number of smokers</i>
$20 \leq x < 30$	5
$30 \leq x < 40$	4
$40 \leq x < 50$	3
$50 \leq x < 60$	2
$60 \leq x < 70$	3

(a) Calculate an estimate of the mean smoking age.

(b) Construct a histogram to represent this data.

4. The following results give the heights of sunflowers in centimetres.

180 184 195 177 175 173 169 167 197 173 166 183 161 195 177
192 161 165

Represent the data by a stem and leaf diagram.

5. The following stem and leaf diagram gives the heights in cm of 39 schoolchildren.

<i>Ste</i>	<i>Leaf</i>	<i>Key</i>
<i>m</i>		<i>2 represents 132</i>
		<i>13 cm.</i>
<i>13</i>	<i>2, 3, 3, 5, 8,</i>	
<i>14</i>	<i>1, 1, 1, 4, 5, 5, 9,</i>	
<i>15</i>	<i>3, 4, 4, 6, 6, 7, 7, 7, 8, 9, 9,</i>	
<i>16</i>	<i>1, 2, 2, 5, 6, 6, 7, 8, 8,</i>	
<i>17</i>	<i>4, 4, 4, 5, 6, 6,</i>	
<i>18</i>	<i>0,</i>	

(a) (i) State the lower quartile height,

(ii) State the median height

(iii) State the upper quartile height.

(b) Draw a box-and-whisker plot for the above information.

References

Research methods by Mugenda Olive M and Mugenda Abel G. Pg126-130

Business Calculations and statistics simplified by N.A Saleemi. Revised Edition. Pg 275-285

Essentials of statistics for Business and Economics by Anderson Sweetey Williams Pg 22-34

LECTURE4

CHAPTER 4: ANALYSIS AND INTERPRETATION OF DATA



Purpose

To examine various measures of central tendency.

Objectives

Compute numerical quantities that measure centrality in a set of data such as:

- i. Arithmetic mean*
- ii. Median*
- iii. Mode*
- iv. Geometric mean*
- v. Weighted mean*
- vi. Harmonic mean*

Measures of Central Tendency

4.1 Introduction

In the previous chapter, we have studied how to collect raw data, its classification and tabulation in a useful form, which contributes in solving many problems of statistical concern. Yet, this is not sufficient, for in practical purposes, there is need for further condensation, particularly when we want to compare two or more different distributions. We may reduce the entire distribution to one number which represents the distribution.

A single value which can be considered as typical or representative of a set of observations and around which the observations can be considered as Centered is called an 'Average' (or average value) or a Centre of location. Since such typical values tend to lie centrally within a set of observations when arranged according to magnitudes, averages are called measures of central tendency.

In fact the distribution have a typical value (average) about which, the observations are more or less symmetrically distributed. This is of great importance, both theoretically and practically. Dr. A.L. Bowley correctly stated, "Statistics may rightly be called the science of averages."

The word average is commonly used in day-to-day conversations. For example, we may say that Okanga is an average boy of my class; we may talk of an average American, average income, etc. When it is said, "Okanga is an average student," it means is that he is neither very good nor very bad, but a mediocre student. However, in statistics the term average has a different meaning.

The fundamental measures of tendencies are:

- (1) Arithmetic mean
- (2) Median
- (3) Mode
- (4) Geometric mean
- (5) Harmonic mean
- (6) Weighted averages

However the most common measures of central tendencies or Locations are Arithmetic mean, median and mode. We therefore, consider the *Arithmetic mean*.

4.2 Arithmetic Mean

This is the most commonly used average which you have also studied and used in lower grades. Here are two definitions given by two great masters of statistics.

Arithmetic mean is the amount secured by dividing the sum of values of the items in a series by their number. Or

The arithmetic average may be defined as the sum of aggregate of a series of items divided by their number.

Thus, you should add all observations (values of all items) together and divide this sum by the number of observations (or items).

4.2.1 Ungrouped Data

Suppose, we have 'n' observations (or measures) $x_1, x_2, x_3, \dots, x_n$ then the Arithmetic mean is

obviously
$$\frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

We shall use the symbol \bar{x} (pronounced as x bar) to denote the Arithmetic mean. Since we have to write the sum of observations very frequently, we use the usual symbol ' \sum ' (pronounced as sigma) to denote the sum. The symbol x_i will be used to denote, in general the 'i' th

observation. Then the sum, $x_1 + x_2 + x_3 + \dots + x_n$ will be represented by $\sum_{i=1}^n x_i$ or $\sum x_i$ simply

Therefore the Arithmetic mean of the set $x_1 + x_2 + x_3 + \dots + x_n$ is given by,

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

This method is known as the "Direct Method".

Example A variable takes the values as given below. Calculate the arithmetic mean of 110, 117, 129, 195, 95, 100, 100, 175, 250 and 750.

Solution: Arithmetic mean =
$$\frac{\sum x_i}{n}$$

$$\sum x_i = 110 + 117 + 129 + 195 + 95 + 100 + 100 + 175 + 250 + 750 = 2021$$

and $n = 10$

Indirect Method (Assumed Mean Method)

$$\bar{u} = \frac{\sum u_i}{n} \text{ Where } u = x_i - A \quad \text{and} \quad A = \text{Assumed Mean}$$

$$\text{Mean} = \bar{x} = A + \bar{u}$$

Calculations:

Let $A = 175$ then

$$u_i = -65, -58, -46, +20, -80, -75, -75, +0, +75, +575$$

$$= 670 - 399$$

$$= 271/10 = 27.1$$

$$\bar{x} = A + \bar{u}$$

$$= 175 + 27.1$$

$$= 202.1$$

Example Mr. Sonko's earnings for the past week were:

Monday	\$ 450
Tuesday	\$ 375
Wednesday	\$ 500
Thursday	\$ 350
Friday	\$ 270

Find his average earning per day.

Solution:

$$\sum_{i=1}^5 x_i = 450 + 375 + 500 + 350 + 270 = \$1945$$
$$n = 5$$

$$\text{Arithmetic_mean} = \frac{\sum x_i}{n} = \frac{1945}{5} = \$389$$

Therefore, Mr. Sonko's average earning per day is \$389.

Short-cut Method :

Sometimes the values of x are very big and in that case, to simplify the calculation the short-cut method is used. For this, first you assume a mean (called as the assumed mean). Let it be A . Now find the deviations of all the values of x from A . We now get a new variable $u_i = x_i - A$

Now find

$$\bar{u} = \frac{\sum u_i}{n} \quad \text{then} \quad \bar{x} = A + \frac{\sum u_i}{n} \quad \text{or} \quad \bar{x} = A + \bar{u}$$

Example The expenditure of ten families in dollars is given below:

Family : A B C D E F G H I J

Expenditure : 300 700 100 750 500 80 120 250 100 370

(in dollars).

Calculate the Arithmetic mean.

Solution: Let the assumed mean be \$ 500. (as. = assume)

Families	Expenditure (\$)	Deviation from as. mean
(x_i)		($u_i = x_i - A$)
A	300	-200
B	700	200
C	100	-400
D	750	250
E	500	0
F	80	-420
G	120	-380
H	250	-250
I	100	-400
J	370	-130
n = 10		$\Sigma u_i = -2180 + 450$ $= -1730$

Calculations :

$$\bar{u} = \frac{\Sigma u_i}{n} = \frac{-1730}{10} = -173$$

$$\bar{x} = A + \bar{u}$$

$$\therefore \bar{x} = 500 + (-173) = 327$$

4.2.2 Grouped data

There is a difference in the methods for finding the arithmetic means of the individual series and a discrete series. In the discrete series, every term (i.e. value of x) is multiplied by its corresponding frequency $f_i x_i$ and then their total (sum) is found $\Sigma f_i x_i$. The arithmetic mean is then obtained by dividing the total frequency Σf by the above sum so obtained $\Sigma f_i x_i$.

Therefore, if the observations $x_1 + x_2 + x_3 + \dots + x_n$ are repeated $f_1 + f_2 + f_3 + \dots + f_n$ times, then we have:

Arithmetic mean can be found by

$$(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i} \text{ i.e. } \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \dots + f_n x_n}{f_1 + f_2 + f_3 + \dots + f_n}$$

The formulae for Arithmetic mean by direct method and by the short-cut methods are as follows:

Direct method

Short-cut method

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} \quad \bar{x} = A + \bar{u} \text{ where } \bar{u} = \frac{\sum f_i u_i}{\sum f_i}$$

and $u = x_i - A$

Therefore,

$$\bar{x} = A + \frac{\sum f_i u_i}{\sum f_i}$$

Example Find the mean of the following 50 observations.

19, 19, 20, 20, 20, 19, 20, 18, 21, 19,

20, 20, 19, 19, 20, 19, 21, 19, 19, 21,

18, 20, 18, 18, 17, 20, 20, 22, 20, 20,

20, 20, 20, 21, 20, 17, 23, 18, 17, 21,

20, 21, 20, 20, 20, 18, 21, 19, 20, 19

Solution: We may tabulate the given observations as follows.

Observations (x_i)	Frequency (f_i)	fx_i
17	3	$17 \times 3 = 51$
18	6	$18 \times 6 = 108$
19	11	$19 \times 11 = 209$
20	20	$20 \times 20 = 400$
21	8	$21 \times 8 = 168$
22	1	$22 \times 1 = 22$
23	1	$23 \times 1 = 23$
Total	$\sum f_i = 50$	$\sum f_i x_i = 981$

The arithmetic mean is

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{981}{50} = 19.62$$

Example Eight coins were tossed together and the number of times they fell on the side of heads was observed. The activity was performed 256 times and the frequency obtained for different values of x , (the number of times it fell on heads) is shown in the following table. Calculate then mean by:

i) Direct method ii) Short-cut method

x : 0 1 2 3 4 5 6 7 8

f : 1 9 26 59 72 52 29 7 1

Solution:

No. of coins	Frequency	Direct method	Short-cut method Dev. from A (4) i.e. $u_i = x_i - A$	
x_i	f_i	$f_i x_i$	u_i	$\Sigma f_i u_i$
0	1	0	-4	-4
1	9	9	-3	-27
2	26	52	-2	-52
3	59	117	-1	-59
4	72	288	0	0
5	52	260	1	+52
6	29	174	2	+58
7	7	49	3	+21
8	1	8	4	+4
	$\Sigma f = 256$	$\Sigma fx = 1017$		+135 - 142 $\Sigma fu = -7$

Example Find the arithmetic mean for the following :

Marks below : 10 20 30 40 50 60 70 80

No. of students : 15 35 60 84 96 127 198 250

Solution:

First, we have to convert the cumulative frequencies into frequencies of the respective classes.

Marks	Mid-values	Frequencies		$U = X - A$ $A = 45$	$f_i u_i$
	x_i	c.f.	f.		
0 - 10	5	15	15	- 40	- 600
10 -20	15	35	20	- 30	- 600
20 - 30	25	60	25	- 20	- 500

30 - 40	35	84	24	- 10	- 240
40 - 50	45	96	12	0	0
50 - 60	55	127	31	+10	+310
60 - 70	65	198	71	+20	+1420
70 - 80	75	250	52	+30	+1560
Total		$\Sigma f = 250$			$\Sigma f_u = 1350$

$$\text{Arithmetic Mean } (\bar{x}) = A + \bar{u}$$

$$= A + \frac{\Sigma f_i u_i}{\Sigma f_i}$$

$$= 45 + \frac{1350}{250}$$

$$= 45 + 5.4$$

$$= 50.4 \text{ marks}$$

Step-Deviation Method

Here all class intervals are of the same width say 'c'. This method is employed in place of the Short-cut method. We measure all the class-marks (mid values) from some convenient value, say 'A', which generally should be taken as the class-mark of a class of maximum frequency or of a class which is the middle one. All the class marks happen to be multiples of c, since all class intervals are equal. We consider class frequencies as if they are centered at the corresponding class-marks.

Theorem If $x_1, x_2, x_3, \dots, x_n$ are n values of the class marks with frequencies $f_1, f_2, f_3, \dots, f_n$ respectively and if each x_i is expressed in terms of the new variable u_i by the relation

$x_i = A + cu_i$ then, with the usual notation, we have $\bar{x} = A + c\bar{u}$

where $\bar{u} = \frac{\sum fu_i}{\sum f_i}$ and $u_i = \frac{x_i - A}{c}$

This method is also known as the "Coding method."

Example Calculate the arithmetic mean from the following data :

Age (years) below : 25 30 35 40 45 50 55 60

No. of employees : 8 23 51 81 103 113 117 120

Solution :

Age (years)	Mid - values x_i	Frequencies		$u_i = \frac{x_i - A}{c}$	fu_i
		c.f	f_i		
20 - 25	22.5	8	8	- 4	- 32
25 - 30	27.5	23	15	- 3	- 45
30 - 35	32.5	51	28	- 2	- 59
35 - 40	37.5	81	30	- 1	- 30
40 - 45	42.5 $\Rightarrow A$	103	22	0	0
45 - 50	47.5	113	10	+1	+ 10
50 - 55	52.5	117	4	+2	+ 8
55 - 60	57.5	120	3	+3	+ 9
Total			$\sum f = 120$		$\sum fu_i = - 136$

Arithmetic mean (\bar{x}) = $A + c\bar{u}$

$$\text{where } \bar{u} = \frac{\sum fu_i}{\sum f_i} = \frac{-136}{120} \text{ and } c = 5$$

$$\therefore \bar{x} = 42.5 + \left(\frac{-136}{120}\right) \times 5$$

$$\therefore \bar{x} = 42.5 - 5.67$$

$$\therefore \bar{x} = 36.83 \text{ years}$$

Example From the following data, of the calculation of arithmetic mean, find the missing item.

wages in : 110 112 113 117 ? 125 129 130

No. of

workers 25 17 13 15 14 8 7 2

Mean wage \$ 115.86

Solution:

Wages in \$ x_i	Number of workers f_i	$f_i x_i$
110	25	2750
112	17	1904
113	13	1496
117	15	1755

x	14	14x
125	8	1000
128	6	768
130	2	260
Total	$\sum f_i = 100$	$\sum f_i x_i = 9906 + 14x$

Now the arithmetic mean $(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i}$

$$\text{Therefore, } 115.86 = \frac{9906 + 14x}{100}$$

$$\text{or } 115.86 \times 100 = 9906 + 14x$$

$$11586 = 9906 + 14x$$

$$11586 - 9906 = 14x$$

$$1680 = 14x$$

$$\text{or } x = \frac{1680}{14}$$

$$x = 120$$

Therefore the missing item is 120.

4.2.3 Properties of Arithmetic Mean

1. The sum of the deviations, of all the values of x, from their arithmetic mean, is zero.

Justification

$$\therefore \sum f_i (x_i - \bar{x}) = \sum f_i x_i - \bar{x} \sum f_i = 0$$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} \therefore \sum f_i x_i = \bar{x} \sum f_i$$

Since \bar{x} is a constant,

2. The product of the arithmetic mean and the number of items gives the total of all items.

Justification :

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} \Rightarrow \sum f_i x_i = \bar{x} \sum f_i$$

$$\text{or } \bar{x} = \frac{\sum x_i}{N} \Rightarrow \bar{x} \cdot N = \sum x_i$$

3. If \bar{x}_1 and \bar{x}_2 are the arithmetic mean of two samples of sizes n_1 and n_2 respectively then, the arithmetic mean \bar{x} of the distribution combining the two can be calculated as

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

This formula can be extended for still more groups or samples.

$$\bar{x}_1 = \frac{\sum x_{1i}}{n_1} \Rightarrow \sum x_{1i} = n_1 \bar{x}_1$$

Justification:

= total of the observations of the first sample

$$\bar{x}_1 = \frac{\sum x_{1i}}{n_1} \Rightarrow \sum x_{1i} = n_1 \bar{x}_1$$

Similarly $\sum x_{2i} = n_2 \bar{x}_2 =$ total of the observations of the first sample

The combined mean of the two samples

$$= \frac{\text{combined total}}{n_1 + n_2}$$

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

Example The average marks of three batches of students having 70, 50 and 30 students respectively are 50, 55 and 45. Find the average marks of all the 150 students, taken together.

Solution :

Let \bar{x} be the average marks of all 150 students taken together.

	Batch - I	Batch - II	Batch - III
	\bar{x}_1	\bar{x}_2	\bar{x}_3
A. marks :	= 50	= 55	= 45
No. of students	$n_1 = 70$	$n_2 = 50$	$n_3 = 30$

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3}{n_1 + n_2 + n_3} = \frac{70 \times 50 + 50 \times 55 + 30 \times 45}{70 + 50 + 30} = \frac{7600}{150}$$

$$\bar{x} = 50.67 \text{ marks}$$

Example The mean of a certain number of observations is 40. If two or more items with values 50 and 64 are added to this data, the mean rises to 42. Find the number of items in the original data.

Solution:

Let 'n' be the number of observations whose mean $\bar{x} = 40$.

$$\bar{x} = \frac{\sum x_i}{n} \therefore \sum x_i = n\bar{x} = n(40) = 40n$$

total of n values.

Two more items of values 50 and 64 are added therefore, total of (n + 2) values :

$$\begin{aligned} &= \sum x_i + 50 + 64 \\ &= 40n + 50 + 64 \\ &= 40n + 114 \end{aligned}$$

Now new mean is 42.

$$\text{New } \bar{x} = \frac{\text{New total of (n + 2) values}}{n + 2}$$

$$42 = \frac{40n + 114}{n + 2}$$

$$42n + 84 = 40n + 114$$

$$2n = 30$$

$$n = 15$$

Therefore, the number of items in the original data = 15.

Example The sum of deviations of a certain numbers of observations measured from 4 is 72 and the sum of deviations of observations measured from 7 is -3. Find the number of observations and their mean.

Solution:

Let 'n' be the required number of observations $\sum (x_i - 4) = 72$, therefore, $\sum x_i - 4n = 72$

.....Note $\sum 4 = 4n$ and $\sum x_i - 7 = -3$

therefore,

$$\sum x_i - 7 = -3$$

Subtracting the two equations we get,

$$\sum x_i - 4n = 72$$

$$\sum x_i - 7n = -3$$

$$\begin{array}{r} (-) \quad (+) \quad (+) \\ \hline \end{array}$$

$$3n = 75$$

$$n = 25$$

Putting $n = 25$ in $\sum x_i - 4n = 72$, we get

$$\sum x_i - 4(25) = 72$$

$$\sum x_i = 72 + 100$$

$$\sum x_i = 172$$

Now Mean is given by $\bar{x} = \frac{\sum x_i}{n} = \frac{172}{25} = 6.88$

Example The mean weight of 98 students is found to be 50 kg. It is later discovered that the frequency of the class interval (30- 40) was wrongly taken as 8 instead of 10. Calculate the correct mean.

Solution:

Incorrect mean $(\bar{x}) = 50 \text{ Kg}$ and $\sum f_i = 98$

$$\text{Incorrect } \bar{x} = \frac{\text{incorrect } \sum f_i x_i}{\sum f_i}$$

$$50 = \frac{\text{incorrect } \sum f_i x_i}{98}$$

Therefore, Incorrect $\sum f_i x_i = 98 \times 50 = 4900$

Now correct

$$\sum f_i x_i = \text{Incorrect } \sum f_i x_i - (8 \times 35) + (10 \times 35)$$

Note that the class-mark of class interval (30 - 40) is 35 and for the calculation of the mean we consider class marks.

$$\text{The correct } \sum f_i x_i = 4900 - 280 + 350 = 4970$$

$$\text{Also the correct } \sum f_i = 98 + 2 = 100$$

Therefore, the correct mean

$$\begin{aligned} \bar{x} &= \frac{\text{correct } \sum f_i x_i}{\text{correct } \sum f_i} = \frac{4970}{100} \\ &= 49.7 \text{kg} \end{aligned}$$

Example The sum of the deviations of 'n' observation values of a variate from a

constant 'a', is S. Show that the arithmetic mean is $a + \frac{S}{n}$.

Solution:

$$\begin{aligned}\sum(x_i - a) &= S \\ \therefore \sum x_i - \sum a &= S \\ \therefore \sum x_i - an &= S \quad \text{since } \sum a = a \sum 1 = a \cdot n\end{aligned}$$

Dividing by n to get the mean

$$\text{i.e. } \frac{\sum x_i}{n} - \frac{an}{n} = \frac{S}{n} \quad \therefore \frac{\sum x_i}{n} = a + \frac{S}{n}$$

$$\text{But } \frac{\sum x_i}{n} = \bar{x}$$

$$\text{Therefore, } \bar{x} = a + \frac{S}{n}$$

LECTURE 5

4.2.4 Merits of the Mean

1. It is rigidly defined. Its value is always definite.
2. It is easy to calculate and easy to understand. Hence it is very popular.
3. It is based on all the observations; so that it becomes a good representative.
4. It can be easily used for comparison.
5. It is capable of further algebraic treatment such as finding the sum of the values of the observations, if the mean and the total number of the observations are given; finding the combined arithmetic mean when different groups are given etc.
6. It is not affected much by sampling fluctuations.

4.2.5 Demerits of the Arithmetic mean

1. It is affected by outliers or extreme values. For example, the average (A.) mean of 10, 15,

25 and 500 is $\bar{x} = \frac{10+15+25+500}{4} = 137.5$

Now observe first three values whose mean(A) is $\frac{10+15+25}{3} = 16.67(\text{approx})$

Due to the outlier 500 the A. mean of the four numbers is raised to 137.5. In such a case A. mean is not a good representative of the given data.

2. It is a value which may not be present in the given data.
3. Many a times it gives absurd results like 4.4 children per family.
4. It is not possible to take out the averages of ratios and percentages.
5. We cannot calculate it when open-end class intervals are present in the data.

4.3 Median

It is the value of the size of the central item of the arranged data (data arranged in the ascending or the descending order). Thus, it is the value of the middle item and divides the series in to equal parts.

In Connor's words - "The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater and the other all values lesser than the median." For example, the daily wages of 7 workers are 5, 7, 9, 11, 12, 14 and 15 dollars. This series contains 7 terms. The fourth term i.e. \$11 is the median.

4.3.1 Median In Individual Series (ungrouped Data)

1. Set the individual series either in the ascending (increasing) or in the descending (decreasing) order, of the size of its items or observations.
2. If the total number of observations be 'n' then
 - A. If 'n' is odd,

$$\text{The median} = \text{size of } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ observation}$$

- B. If 'n' is even, the median

$$= \frac{1}{2} \left[\begin{array}{l} \text{size of } \left(\frac{n}{2}\right)^{\text{th}} \text{ observations} \\ + \text{size of } \left(\frac{n+2}{2}\right)^{\text{th}} \text{ observations} \end{array} \right]$$

Example The following figures represent the number of books issued at the counter of a Statistics library on 11 different days. 96, 180, 98, 75, 270, 80, 102, 100, 94, 75 and 200. Calculate the median.

Solution:

Arrange the data in the ascending order as 75, 75, 80, 94, 96, 98, 100, 102, 180, 200, 270.

Now the total number of items 'n'= 11 (odd)

Therefore, the median = size of $\left(\frac{n+1}{2}\right)^{\text{th}}$ item

$$\begin{aligned} &= \text{size of } \left(\frac{11+1}{2}\right)^{\text{th}} \text{ item} \\ &= \text{size of } 5^{\text{th}} \text{ item} \\ &= 98 \text{ books per day} \end{aligned}$$

Example The population (in thousands) of 36 metropolitan cities are as follows :

2468, 591, 437, 20, 213, 143, 1490, 407, 284, 176, 263, 19, 181, 777, 387, 302, 213, 204, 153, 733, 391, 176 178, 122, 532, 360, 65, 260, 193, 92, 672, 258, 239, 160, 147, 151. Calculate the median.

Solution:

Arranging the terms in the ascending order as:

20, 65, 92, 131, 142, 143, 147, 151, 153, 160, 169, 176, 178, 181, 193, 204, (213, 39), 258, 263, 260, 384, 302, 360, 387, 391, 407, 437, 522, 591, 672, 733, 777, 1490, 2488.

Since total number of items n = 36 (Even).

the median

$$= \frac{1}{2} \left[\text{size of } \left(\frac{n}{2}\right)^{\text{th}} \text{ item} + \text{size of } \left(\frac{n+2}{2}\right)^{\text{th}} \text{ item} \right]$$

$$\begin{aligned}
&= \frac{1}{2} [\text{size of } 18^{\text{th}} \text{ item} + \text{size of } 19^{\text{th}} \text{ item}] \\
&= \frac{1}{2} [213 + 239] \\
&= \frac{1}{2} [252] \\
&= 226 \text{ thousands}
\end{aligned}$$

4.3.2 Median In Discrete Series

Steps :

1. Arrange the data in ascending or descending order of magnitude.
2. Find the cumulative frequencies.
3. Apply the formula :

A. If 'n' = $\sum f_i$ (odd) then,

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

B. If 'n' = $\sum f_i$ (even) then,

$$\text{Median} = \frac{1}{2} \left[\text{size of } \left(\frac{n}{2} \right)^{\text{th}} \text{ item} + \text{size of } \left(\frac{n+2}{2} \right)^{\text{th}} \text{ item} \right]$$

Example Locate the median in the following distribution.

Size : 8 10 12 14 16 18 20

Frequency : 7 7 12 28 10 9 6

Solution:

Size (x_i)	Frequency f_i	Cumulative frequency (c.f)
8	3	3
10	7	$3 + 7 = 10$
12	12	$12 + 10 = 22$
14	28	$28 + 22 = 50$
16	10	$50 + 10 = 60$
18	9	$60 + 9 = 69$
20	6	$69 + 6 = 75$
	$n = \sum f_i = 75$ (odd)	

Therefore, the median = size of $\left(\frac{n+1}{2}\right)^{\text{th}}$ item

$$= \text{size of } \left(\frac{75+1}{2}\right)^{\text{th}} \text{ item}$$

= size of 38th item

In the order of the cumulative frequency, the 38th term is present in the 50th cumulative frequency, whose size is 14.

Therefore, the median = 14

4.3.3 Median in Continuous Series (grouped Data)

Steps:

1. Determine the particular class in which the value of the median lies. Use $n/2$ as the rank

of the median and not $\left(\frac{n+1}{2}\right)$

2. After ascertaining the class in which median lies, the following formula is used for determining the exact value of the median.

$$\text{Median} = \ell_1 + \left[\frac{N/2 - \text{c.f}}{f} \right] (\ell_2 - \ell_1)$$

where,

ℓ_1 = lower limit of the median class, the class in which the middle item of the distribution lies.

ℓ_2 = upper limit of the median class

c.f = cumulative frequency of the class preceding the median class

f = sample frequency of the median class

It should be noted that while interpolating the median value of frequency distribution it is assumed that the variable is continuous and that there is an orderly and even distribution of items within each class.

Example Calculate the median for the following and verify it graphically.

Age (years) : 20-25 25-30 30-35 35-40 40-45

No. of person : 70 80 180 150 20

Solution:

Age (years) C.I.	No. of persons f_i	Cumulative frequency c.f.
20 - 25	70	70
25 - 30	80	$70 + 80 = 150$
30 - 35	180	$150 + 180 = 330$
35 - 40	150	$330 + 150 = 480$
40 - 45	20	$480 + 20 = 500$
	$n = \sum f_i = 500$	

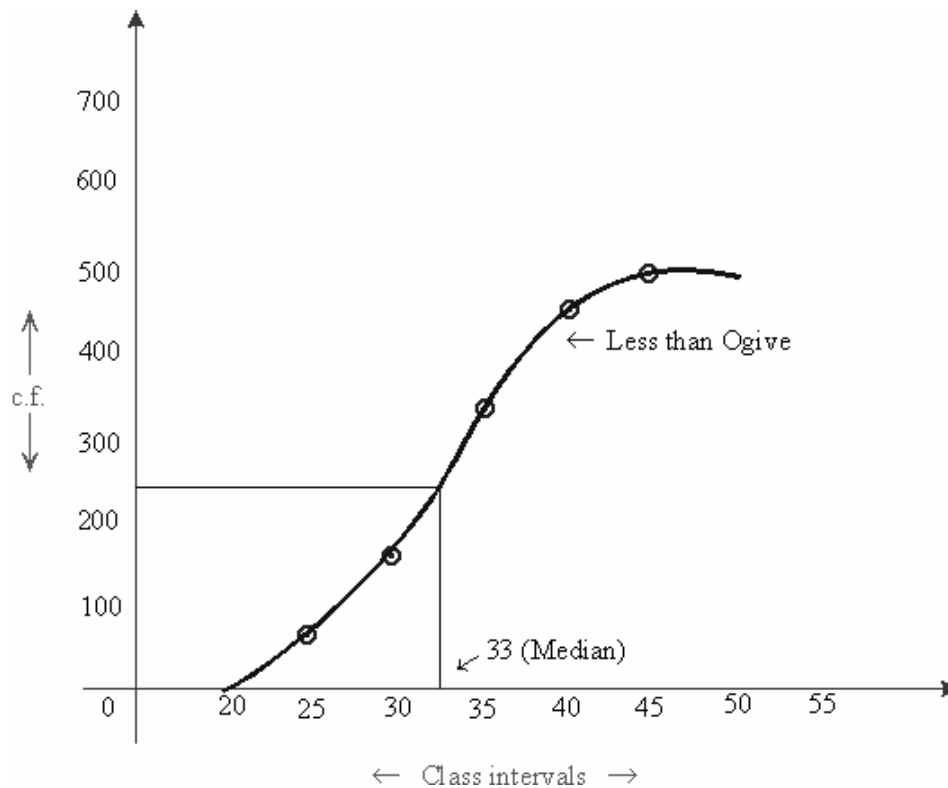
Now median = size of $(n/2)^{\text{th}}$ item
= size of $\left(\frac{500}{2}\right)^{\text{th}}$
= size of 250^{th} item which
lies in (30 - 35) class interval

$$\text{Median} = \ell_1 + \left[\frac{N/2 - \text{c.f.}}{f} \right] (\ell_2 - \ell_1)$$

Here $\ell_1 = 30$, $\ell_2 = 35$, $n/2 = 250$, c.f. = 150 and $f = 180$

Therefore, Median

$$\begin{aligned} &= 30 + \left[\frac{250 - 150}{180} \right] (35 - 30) \\ &= 30 + \frac{100}{180} \times 5 \\ &= 32.78 \text{ years} \end{aligned}$$



Sometimes the series is given in the descending order of magnitude. In this situation convert the series in the ascending order of magnitude and then using the regular formula, the median can be calculated or the series can be put in the descending order of the magnitude and an alternative formula be used to calculate the median

Example Marks : 40 -50 30- 40 20-30 10-20 0 -10

No. of students : 10 12 40 30 8

Solution :

Marks	No. of students	c.f.
0 -10	8	8
10 - 20	30	8 + 30 = 38
20 - 30	40	38 + 40 = 78
30 - 40	12	78 + 12 = 90
40 - 50	10	90 + 10 = 100
	$n = \sum f = 100$	

Median = size of $(n/2)^{\text{th}}$ item

$$= \text{size of } \left(\frac{100}{2}\right)^{\text{th}} \text{ item}$$

= size of 50^{th} item which comes

into class interval (20 – 30)

By interpolation

$$= 20 + \left[\frac{50-38}{40}\right] (30-20)$$

$$= 20 + \frac{12}{40} \times 10$$

$$= 20 + 3$$

$$\text{Median} = \ell_1 + \left[\frac{n/2 - \text{c.f.}}{f}\right] (\ell_2 - \ell_1) \therefore \text{Median} = 23 \text{ marks}$$

Arranging the series in the descending order (as it is given)

Marks	No. of students	Cumulative Frequency
40 - 50	10	10
30 - 40	12	22
20 - 30	40	62
10 - 20	30	92
0 - 10	8	100
	$n = \sum f = 100$	

Median = size of $\left(\frac{n}{2}\right)^{th}$ item = size of 50th item which lies in (20 -30) class-interval.

Alternative formula :

Median

$$\begin{aligned}
 &= l_2 - \left[\frac{n/2 - c.f}{f} \right] (l_2 - l_1) \\
 &= 30 - \left[\frac{50 - 22}{40} \right] (30 - 20) \\
 &= 30 - \frac{28}{40} \times 10 \\
 &= 30 - 7 \\
 &= 23 \text{ marks}
 \end{aligned}$$

Note that, while calculating the median of a series, it must be put in the 'exclusive class-interval' form. If the original series is in inclusive type, first convert it into the exclusive type and then find its median.

Example The following distribution represents the number of minutes spent by a group of teenagers in watching movies. What is the median?

Minutes/Weeks: 0-99 100-199 200-299 300-399 400 - 499 500 - 599 600 & more

No. of teenagers: 27 32 65 78 58 32 8

Solution:

Minutes/ weeks	Real class intervals	Frequency	Cumulative frequency
0 -99	0.5 - 99.5	27	27
100 - 199	99.5 - 199.5	32	27 + 32 = 59
200 - 299	199.5 - 299.5	65	59 + 65 = 124
300 - 399	299.5 - 399.5	78	124 + 78 = 202
400 - 499	399.5 - 499.5	58	202 + 58 = 260
500 - 599	499.5 - 599.5	32	260 + 32 = 292
600 & more	599.5 & more	82	292 + 8 = 300
		$n = \sum f_i = 300$	

Median = size of $(n/2)^{\text{th}}$ item
 = size of $\left(\frac{300}{2}\right)^{\text{th}}$ item
 = size of 150^{th} item which
 lies in (299.5 – 399.5) class interval

By using interpolation

$$\begin{aligned} \text{Median} &= \ell_1 - \left[\frac{n/2 - c.f.}{f} \right] (\ell_2 - \ell_1) \\ &= 299.5 + \left[\frac{150 - 124}{78} \right] (399.5 - 299.5) \\ &= 299.5 + \left[\frac{26}{78} \right] \times 100 \\ &= 299.5 + 33.33 \text{ (approximately)} \\ \text{Median} &= 332.88 \text{ minutes/ week} \end{aligned}$$

LECTURE 6

4.3.4 Merits of Median

1. It is rigidly defined.
2. It is easy to calculate and understand.
3. It is not affected by extreme values like the arithmetic mean. For example, 5 persons have their incomes \$2000, \$2500, \$2600, \$3000, \$5000. The median would be \$2600 while the arithmetic mean would be \$3020.
4. It can be found by mere inspection.
5. It is fully representative and can be computed easily.
6. It can be used for qualitative studies.
7. Even if the extreme values are unknown, median can be calculated if one knows the number of items.
8. It can be obtained graphically.

4.3.5 Demerits of Median

1. It may not be representative if the distribution is irregular and abnormal.
2. It is not capable of further algebraic treatment.
3. It is not based on all observations.
4. It is affected by sample fluctuations.
5. The arrangement of the data in the order of magnitude is absolutely necessary.

4.4 Mode

It is the size of that item which possesses the maximum frequency. According to Professor Kenney and Keeping, the value of the variable which occurs most frequently in a distribution is called the mode.

It is the most common value. It is the point of maximum density.

4.4.1 Ungrouped Data

Individual series : The mode of this series can be obtained by mere inspection. The number which occurs most often is the mode.

Example Locate mode in the data 7, 12, 8, 5, 9, 6, 10, 9, 4, 9, 9

Solution : On inspection, it is observed that the number 9 has maximum frequency. Therefore 9 is the mode.

Note that if in any series, two or more numbers have the maximum frequency, then the mode will be difficult to calculate. Such series are called as Bi-modal, Tri-modal or Multi-modal series.

4.4.2 Grouped Data

Steps :

1. Determine the modal class which has the maximum frequency.
2. By interpolation the value of the mode can be calculated as -

$$\text{Mode} = \ell_1 - \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right] (\ell_2 - \ell_1)$$

where

ℓ_1 = lower limit of the modal class

ℓ_2 = upper limit of the modal class

f_1 = frequency of the modal class

f_0 = frequency of the class preceding to the modal class

f_2 = frequency of the class succeeding the modal class

Example Calculate the modal wages.

Daily wages in \$: 20 -25 25-30 30-35 35-40 40-45 45-50

No. of workers : 1 3 8 12 7 5

Verify it graphically.

Solution:

Here the maximum frequency is 12, corresponding to the class interval (35 - 40) which is the modal class.

Therefore $\ell_1 = 35$, $\ell_2 = 40$, $f_0 = 8$, $f_1 = 12$, $f_2 = 7$

By interpolation

$$\text{Mode} = \ell_1 - \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right] (\ell_2 - \ell_1)$$

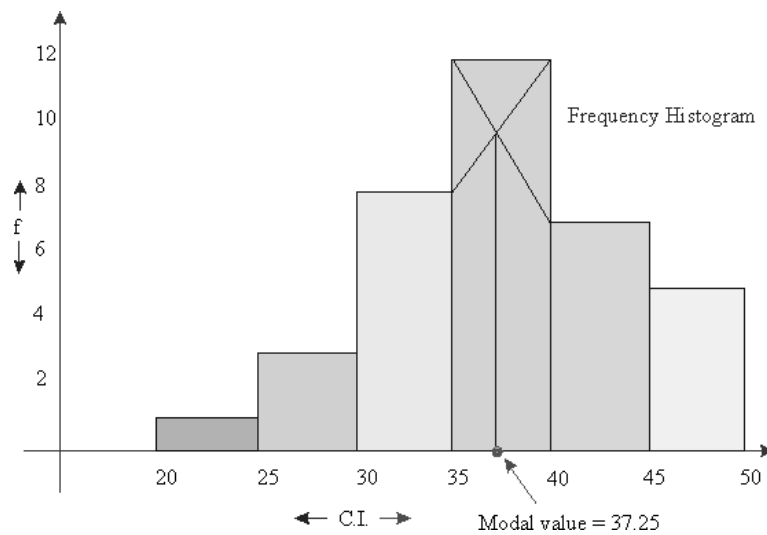
$$= 40 + \left[\frac{12 - 8}{24 - 8 - 7} \right] (40 - 35)$$

$$= 40 + \left[\frac{4}{9} \right] \times 5$$

$$= 40 + 2.22$$

$$= 37.22$$

Modal wages is \$37.22



4.4.3 Merits of Mode

1. It is simple to calculate.
2. In individual or discrete distribution it can be located by mere inspection.
3. It is easy to understand. Everyone is used to the idea of average size of a garment, an average American etc.
4. It is not isolated like the median as it is the most common item.
5. Like the Average mean, it is not a value which cannot be found in the series.
6. It is not necessary to know all the items. What we need the point of maximum density frequency.
7. It is not affected by sampling fluctuations.

4.4.4 Demerits of Mode

1. It is ill defined.
2. It is not based on all observations.
3. It is not capable of further algebraic treatment.
4. It is not a good representative of the data.
5. Sometimes there are more than one values of mode.

4.5 Empirical Relation Between Mean, Median And Mode

A distribution in which the values of mean, median and mode coincide (i.e. mean = median = mode) is known as a symmetrical distribution. Conversely, when values of mean, median and mode are not equal the distribution is known as asymmetrical or skewed distribution. In moderately skewed or asymmetrical distribution a very important relationship exists among these three measures of central tendency. In such distributions the distance between the mean and median is about one-third of the distance between the mean and mode, as will be clear from the diagrams 1 and 2. Karl Pearson expressed this relationship as:

$$\text{Mode} = \text{mean} - 3 [\text{mean} - \text{median}]$$

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

$$\text{and Median} = \text{mode} + \frac{2}{3}[\text{mean} - \text{mode}]$$

Knowing any two values, the third can be computed.

Example

Given median = 20.6, mode = 26 Find mean.

Solution:

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

$$\therefore \text{Mean} = \frac{1}{2}[3 \text{ median} - \text{mode}]$$

$$\therefore \text{Mean} = \frac{1}{2}[3(20.6) - (26)]$$

$$\therefore \text{Mean} = \frac{1}{2}[35.8]$$

$$\therefore \text{Mean} = 17.9$$

4.6 Weighted mean

The weighted mean is a mean where there is some variation in the relative contribution of individual data values to the mean. Each data value (X_i) has a weight assigned to it (W_i). Data values with larger weights contribute more to the weighted mean and data values with smaller weights contribute less to the weighted mean. The formula is

$$\bar{X}_w = \frac{\sum W_i X_i}{\sum W_i}$$

There are several reasons why you might want to use a weighted mean.

1. Each individual data value might actually represent a value that is used by multiple people in your sample. The weight, then, is the number of people associated with that particular value.
2. Your sample might deliberately over represent or under represent certain segments of the population. To restore balance, you would place less weight on the over represented segments of the population and greater weight on the underrepresented segments of the population.
3. Some values in your data sample might be known to be more variable (less precise) than other values. You would place greater weight on those data values known to have greater precision.

Example

Joan gets quiz grades of 79, 82, and 69. She gets a 65 on her final exam. Find the weighted mean if the quizzes each count for 10% and the final exam counts for 70% of the final grade.

Solution

The information can be organized as follows:

X_i	W_i	$W_i X_i$
79	10	790
82	10	820
69	10	690
65	70	4550
	100	6850

$$\bar{X}_w = \frac{\sum W_i X_i}{\sum W_i}$$

$$\begin{aligned}\overline{X}_w &= \frac{6850}{100} \\ &= 68.5\%\end{aligned}$$

4.7 Geometric mean

The geometric mean is an average calculated by multiplying a set of numbers and taking the n th root, where n is the number of numbers.

A common example when the geometric mean is use is when averaging growth rates.

The formula for the geometric mean:-

$$G.M = \sqrt[n]{([x_1][x_2][x_3] \dots [x_n])}$$

Where n is the number of observations made of the variable x and X_1, X_2, \dots, X_n are the values of these observations

For example, the geometric mean of numbers: 3, 25 and 45

There are three observation, thus $n = 3$

$$\begin{aligned}G.M &= \sqrt[3]{([3][25][45])} \\ &= 15\end{aligned}$$

The geometric mean cannot be calculated if we have negative or zero observations. The geometric mean of a set of readings is always less than the arithmetic mean (unless all readings are identical) and is less influenced by very large values / items.

Take the arithmetic mean of the following salaries: - in thousands of shillings per month

6, 8, 10, 10,10,12,16

Arithmetic mean = Kshs 10, 286 per month to the nearest shilling. The geometric men of the salaries are:-

$$\begin{aligned}G.M &= \sqrt[7]{([6][8][10][10][10][12][16])} \\ &= \sqrt[7]{9216000}\end{aligned}$$

$$= 9.884 \text{ (to 3 decimal places)}$$

The geometric mean of the salaries is Kshs.9884 per month

(To the nearest shilling)

Given the following salaries (i.e. in thousands of Kshs) in accompany per annum (p.a):-

6, 8, 10, 10,10,12,48

The arithmetic mean salary is Kshs 14,857 to the nearest shilling. The geometric mean is:-

$$G.M = \sqrt[7]{([6][8][10][10][10][12][48])}$$
$$= \sqrt[7]{27648000}$$

11.564

11.564 to three decimal places

The geometric mean salary is Kshs.11.564 per annum to the nearest shilling. The geometric mean is useful when only a few items in a distribution are changing: it's in the circumstances more stable than the arithmetic mean. It is useful in the calculation of share indices and also in such calculations where data grows in geometric progression i.e. the population of a country

Given population in a city 300,000 in 1980 and 400,000 in 1990, if we wanted to find out an estimate of the arithmetic mean of the population in 1985.

$$\frac{300,000 + 400,000}{2} = \frac{700,000}{2}$$
$$= 350,00$$

Here, we are making an assumption the population grows by the same number each year which is not correct. The same thing applies to money assuming its growing in a compound rate. The geometric mean for 1985 would be:-

$$= \sqrt[2]{300,000 \times 400,000}$$
$$= 371,080$$

4.8 Harmonic mean

Harmonic mean is another measure of central tendency and also based on mathematic footing like arithmetic mean and geometric mean. Like arithmetic mean and geometric mean, harmonic

mean is also useful for quantitative data. Harmonic mean is defined in following terms:

Harmonic mean is quotient of “number of the given values” and “sum of the reciprocals of the given values”.

Harmonic mean in mathematical terms is defined as follows:

For Ungrouped Data	For Grouped Data
$H.M = \bar{X} = \frac{n}{\sum\left(\frac{1}{x}\right)}$	$HM = \bar{X} = \frac{\sum f}{\sum\left(\frac{f}{x}\right)}$

Example:

Calculate the harmonic mean of the numbers: 13.5, 14.5, 14.8, 15.2 and 16.1

Solution:

The harmonic mean is calculated as below:

x	$\frac{1}{x}$
13.2	0.0758
14.2	0.0704
14.8	0.0676
15.2	0.0658
16.1	0.0621
Total	$\sum\left(\frac{1}{x}\right) = 0.3417$

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

$$H.M = \bar{X} = \frac{5}{0.3417} = 14.63$$

Example:

Given the following frequency distribution of first year students of a particular college. Calculate the Harmonic Mean.

Age (Years)	13	14	15	16	17
Number of Students	2	5	13	7	3

Solution:

The given distribution belongs to a grouped data and the variable involved is ages of first year students. While the number of students Represent frequencies.

Ages (Years) x	Number of Students f	$\frac{1}{x}$
13	2	0.1538
14	5	0.3571
15	13	0.8667
16	7	0.4375
17	3	0.1765
Total	$\sum f = 30$	$\sum \frac{1}{x} = 1.9916$

Now we will find the Harmonic Mean as

$$HM = \bar{X} = \frac{\sum f}{\sum \left(\frac{f}{x}\right)} = \frac{30}{1.9916} = 15.0631 \approx 15$$

Example:

Calculate the harmonic mean for the given below:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
F	2	3	11	20	32	25	7

Solution:

The necessary calculations are given below:

Marks	X	F	$\frac{f}{x}$
30-39	34.5	2	0.0580
40-49	44.5	3	0.0674
50-59	54.5	11	0.2018
60-69	64.5	20	0.3101
70-79	74.5	32	0.4295
80-89	84.5	25	0.2959
90-99	94.5	7	0.0741
Total		$\sum f = 100$	$\sum \frac{f}{x} = 1.4368$

Now we will find the Harmonic Mean as

$$\bar{x} = \frac{\sum f}{\sum \left(\frac{f}{x}\right)} = \frac{100}{1.4368} = 69.60$$

4.9 Characteristics of a Good Measure of Central Tendency

1. It should be rigidly defined
2. It should be easy to understand and calculate
3. It should be based on all observations
4. It should be amenable to further algebraic manipulation.
5. It should not be affected much by extreme values
6. It should be least affected by fluctuations in sampling.

Chapter Review Questions

1. *The mean of the ten numbers listed below is 5.5.*

4, 3, a, 8, 7, 3, 9, 5, 8, 3

(a) Find the value of a.

(b) Find the median of these numbers.

2. *In the following ordered data, the mean is 6 and the median is 5.*

2, b, 3, a, 6, 9, 10, 12

Find each of the following

(a) the value of a;

(b) the value of b.

3. *For the set of {8, 4, 2, 10, 2, 5, 9, 12, 2, 6}*

(a) calculate the mean;

(b) find the mode;

(c) find the median.

4. David looked at a passage from a book. He recorded the number of words in each sentence as shown in the following frequency table.

<i>Class interval (number of words)</i>	<i>Frequency f</i>
<i>1–5</i>	<i>16</i>
<i>6–10</i>	<i>28</i>
<i>11–15</i>	<i>26</i>
<i>16–20</i>	<i>14</i>
<i>21–25</i>	<i>10</i>
<i>26–30</i>	<i>3</i>
<i>31–35</i>	<i>1</i>
<i>36–40</i>	<i>0</i>
<i>41–45</i>	<i>2</i>

- (a) Find the class interval in which the median lies.
- (b) Estimate, correct to the nearest whole number, the mean number of words in a sentence.

5. Twenty students are asked how many detentions they received during the previous week at school. The results are summarized in the frequency distribution table below.

<i>Number of detentions x</i>	<i>Number of students f</i>	<i>fx</i>
<i>0</i>	<i>6</i>	
<i>1</i>	<i>3</i>	

<i>2</i>	<i>10</i>	
<i>3</i>	<i>1</i>	
<i>Total</i>	<i>20</i>	

(a) What is the modal number of detentions received?

6. The weight in kilograms of 12 students in a class are as follows.

63 76 99 65 63 51 52 95 63 71 65 83

(a) State the mode.

(b) Calculate the mean weight;

When one student leaves the class, the mean weight of the remaining 11 students becomes 70 kg.

(c) Find the weight of the student who left.

7. An atlas gives the following information about the approximate population of some cities in the year 2000. The population of Nairobi has accidentally been left out.

<i>City</i>	<i>Population in Millions</i>
<i>Melbourne</i>	<i>3.2</i>
<i>Bangkok</i>	<i>7.2</i>
<i>Nairobi</i>	
<i>Paris</i>	<i>9.6</i>

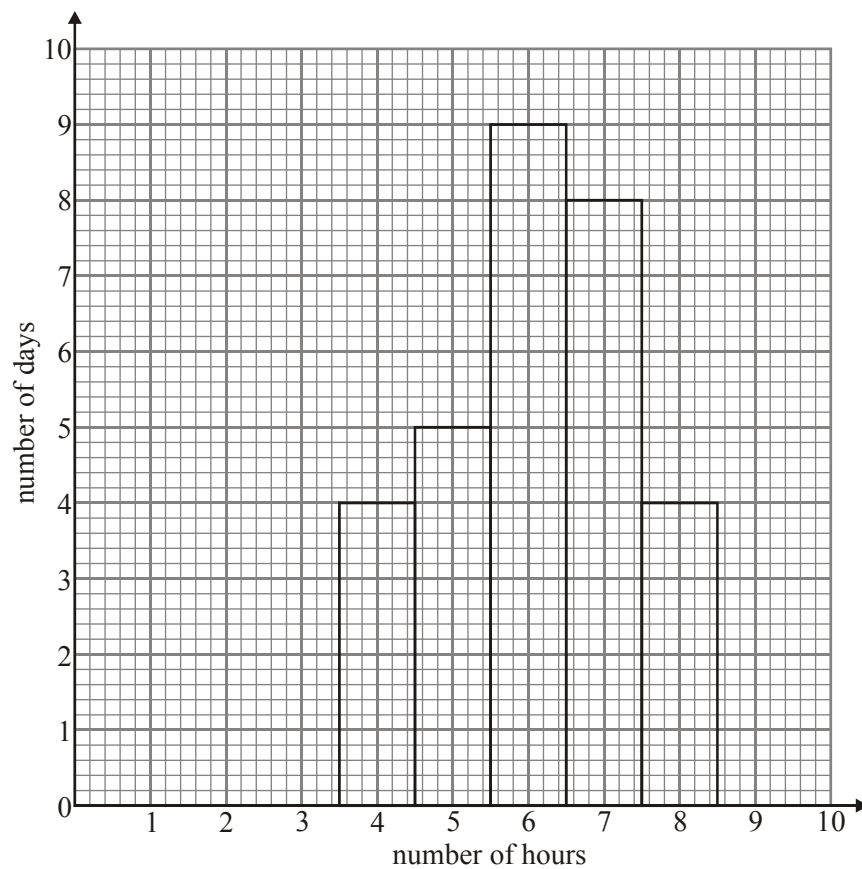
<i>São Paulo</i>	<i>17.7</i>
<i>Tokyo</i>	<i>28.0</i>
<i>Seattle</i>	<i>2.1</i>

The atlas tells us that the mean population for this group of cities is 10.01 million.

(a) Calculate the population of Nairobi.

(b) Which city has the median population value?

8. The number of hours that a professional footballer trains each day in the month of June is represented in the following histogram.



(a) Write down the modal number of hours trained each day.

(b) Calculate the mean number of hours he trains each day.

9. *The numbers of games played in each set of a tennis tournament were*

*9, 7, 8, 11, 9, 6, 10, 8, 12, 6, 8, 13, 7, 9, 10, 9, 10, 11,
12, 8, 7, 13, 10, 7, 7.*

The raw data has been organised in the frequency table below.

<i>games</i>	<i>frequency</i>
<i>6</i>	<i>2</i>
<i>7</i>	<i>5</i>
<i>8</i>	<i>n</i>
<i>9</i>	<i>4</i>
<i>10</i>	<i>4</i>
<i>11</i>	<i>2</i>
<i>12</i>	<i>2</i>
<i>13</i>	<i>2</i>

- (a) Write down the value of n.*
- (b) Calculate the mean number of games played per set.*
- (c) What percentage of the sets had more than 10 games?*
- (d) What is the modal number of games?*

References

Business Calculations and statistics simplified by N.A Saleemi. Revised Edition. Pg 318-365

Essentials of statistics for Business and Economics by Anderson Sweetey Williams Pg 61-67

LECTURE 7

CHAPTER 5: MEASURES OF DISPERSION



Purpose

To introduce the learner to various measures of disparity in a given set of data.

Objectives

- a) Define a measure of dispersion and differentiate it from a measure of central tendency.*
- b) Compute various numerical quantities that measure dispersion.*
- c) Define skewness and kurtosis and compute skewness.*
- d) Explain properties of a good measure of dispersion.*
- e) Apply these measures in summarizing a business environment.*

5.1 Introduction

The measures of central tendencies (i.e. means) indicate the general magnitude of the data and locate only the center of a distribution of measures. They do not establish the degree of variability or the spread out or scatter of the individual items and their deviation from (or the difference with) the means.

i) According to Ncswanger, "Two distributions of statistical data may be symmetrical and have common means, medians and modes and identical frequencies in the modal class. Yet with these points in common they may differ widely in the scatter or in their values about the measures of central tendencies."

ii) Simpson and Kafka said, "An average alone does not tell the full story. It is hardly fully representative of a mass, unless we know the manner in which the individual item. Scatter around it a further description of a series is necessary, if we are to gauge how representative the average is."

From this discussion we now focus our attention on the scatter or variability which is known as **dispersion**. Let us take the following three sets.

Students	Group X	Group Y	Group Z
1	50	45	30
2	50	50	45
3	50	55	75
mean \bar{x}	50	50	50

Thus, the three groups have same mean i.e. 50. In fact the median of group X and Y are also equal. Now if one would say that the students from the three groups are of equal capabilities, it is totally a wrong conclusion then. Close examination reveals that in group X students have equal marks as the mean, students from group Y are very close to the mean but in the third group Z, the marks are widely scattered. It is thus clear that the measures of the central tendency is alone not sufficient to describe the data.

Definition of dispersion: The arithmetic mean of the deviations of the values of the individual items from the measure of a particular central tendency used. Thus the 'dispersion' is also known as the "**average of the second degree.**" Prof. Griffin and Dr. Bowley said the same about the dispersion.

In measuring dispersion, it is imperative to know the amount of variation (absolute measure) and the degree of variation (relative measure). In the former case we consider the range, mean deviation, standard deviation etc. In the latter case we consider the coefficient of range, the coefficient mean deviation, the coefficient of variation etc.

5.2 Methods of Computing Dispersion

Note that, we are going to study some of these and not all.

5.2.1 Range

In any statistical series, the difference between the largest and the smallest values is called as the range.

Thus Range (R) = L - S $\left\{ \begin{array}{l} \mathbf{L} = \text{Largest value of the series.} \\ \mathbf{S} = \text{Smallest value of the series} \end{array} \right.$

Coefficient of Range : The relative measure of the range. It is used in the comparative study of

the dispersion co-efficient of Range = $\frac{L - S}{L + S}$

Example (Individual series) Find the range and the co-efficient of the range of the following items :

110, 117, 129, 197, 190, 100, 100, 178, 255, 790.

Solution: R = L - S = 790 - 100 = 690

Co-efficient of Range = $\frac{L - S}{L + S} = \frac{790 - 100}{790 + 100} = \frac{690}{890} = 0.78$

5.2.2 Mean Deviation

The **mean deviation** of a statistical data is defined as the arithmetic mean of the numerical values of the deviations of items from some average value.

Mean deviation is also known as **average deviation**.

The mean deviation is generally denoted by M.D.

Formula for mean deviation for ungrouped data or an individual series is given by

$MD = \frac{\sum|x - a|}{n}$, where a is the average about which M.D is to be calculated.

For a frequency distribution, the formula for mean deviation is given by

$MD = \frac{\sum f|x - a|}{N}$, where a is the average about which M.D is to be calculated.

$$\text{Coeff. of M.D} = \frac{\text{M.D}}{\text{A.M}}$$

where M.D is the Mean Deviation and A.M is the Arithmetic Mean.

If M.D is calculated about mean, then it is written as M.D (\bar{x}).

$$MD(\bar{x}) = \frac{\sum|x - \bar{x}|}{n} \quad (\text{For individual series})$$

$$MD(\bar{x}) = \frac{\sum f|x - \bar{x}|}{N} \quad (\text{For frequency distribution})$$

$$\text{Coeff. of MD.}(\bar{x}) = \frac{MD.(\bar{x})}{\bar{x}}$$

Example 1:

Find the mean deviation from the mean for the given raw data

12 ,6, 7, 3, 15, 10, 18, 5.

Solution

$$\text{Arithmetic mean} = \bar{x} = \frac{12 + 6 + 7 + 3 + 15 + 10 + 18 + 5}{8}$$

$$= \frac{76}{8} = 9.5$$

x_i	$ x_i - \bar{x} $
12	2.5
6	3.5
7	2.5
3	6.5
15	5.5
10	0.5
18	8.5
5	4.5
	34

$$\text{Mean deviation from the mean} = \frac{34}{8} = 4.25$$

Example (Continuous series) calculate the mean deviation and the coefficient of mean deviation from the following data using the mean.

Difference in ages between boys and girls of a class.

Diff. in years:	No. of students:
0 – 5	449

5 – 10	705
10 – 15	507
15 – 20	281
20 – 25	109
25 – 30	52
30 – 35	16
35 – 40	4

Solution :

Diff. in age	Mid-values (x_i)	frequency (f_i)	$f_i x_i$	$ x_i - \bar{x} $	$f_i x_i - \bar{x} $
0 - 5	2.5	449	1122.5	8	3592
5 - 10	7.5	705	5287.5	3	2115
10 - 15	12.5	507	6337.5	2	1014
15 - 20	17.5	281	4917.5	7	1967
20 - 25	22.5	109	2452.5	12	1308
25 - 30	27.5	52	1430.0	17	884
30 - 35	32.5	16	520.0	22	352
35 - 40	37.5	4	150.0	27	108
		$n=2123$	$\Sigma f_i x_i = 22217.5$		$\Sigma f_i x_i - \bar{x} = 11440$

Calculation:

$$1) \bar{x} = \frac{\sum f_i x_i}{n} = \frac{22217.5}{2123} = 10.5(\text{approx})$$

2) M. D.

$$= \frac{\sum f_i |x_i - \bar{x}|}{n} = \frac{11440}{2123} = 5.4$$

3) Co efficient of M. D.

$$= \frac{\text{M.D.}}{\bar{x}} = \frac{5.4}{10.5} = 0.514$$

5.2.3 Variance

The term variance was used to describe the square of the standard deviation. The concept of variance is of great importance in advanced work where it is possible to split the total into several parts, each attributable to one of the factors causing variations in their original series. Variance is defined as follows:

$$\text{var} = \frac{\sum (x_i - \bar{x})^2}{n}$$

5.2.4 Standard Deviation (s. d.)

It is the square root of the arithmetic mean of the square deviations of various values from their arithmetic mean. it is denoted by s.d (when dealing with a sample) or σ (when dealing with the population)

$$\text{Thus } s.d(\sigma) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \text{ for ungrouped data}$$

$$= s.d = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n}} \text{ for grouped data}$$

$$\text{where } n = \sum f_i$$

Merits :

- (1) It is rigidly defined and based on all observations.
- (2) It is amenable to further algebraic treatment.
- (3) It is not affected by sampling fluctuations.
- (4) It is less erratic

Demerits :

- (1) It is difficult to understand and calculate.
- (2) It gives greater weight to extreme values.

5.3 Co-efficient Of Variation (C. V.)

To compare the variations (dispersion) of two different series, relative measures of standard deviation must be calculated. This is known as co-efficient of variation or the co-efficient of s. d. Its formula is

$$C. V. = \frac{\sigma_x}{x} \times 100$$

Thus it is defined as the ratio s. d. to its mean.

Remark: It is given as a percentage and is used to compare the consistency or variability of two more series. The higher the C. V., the higher the variability and lower the C. V., the higher is the consistency of the data.

Example Calculate the standard deviation and its co-efficient from the following data.

A	B	C	D	E	F	G	H	I	J
10	12	16	8	25	30	14	11	13	11

Solution :

No.	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
A	10	-5	25
B	12	-3	9
C	16	+1	1
D	8	-7	49
E	25	+10	100

F	30	+15	225
G	14	-1	1
H	11	-5	16
I	13	-2	4
J	11	-4	16
n=	$\sum x_i =$		$\sum x_i - \bar{x} ^2 =$
10	150		446

Calculations :

i) $\bar{x} = \frac{\sum x_i}{n} = \frac{150}{10} = 15$

ii) $s.d.(\sigma_x) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{446}{10}} = 6.7$

iii) $\text{co-efficient of s.d.} = \frac{\sigma_x}{\bar{x}} = \frac{6.7}{15} = 0.45$

= 45%

Example Calculate s.d. of the marks of 100 students.

Marks	No. of students	Mid-values	$f_i x_i$	$f_i x_i^2$

	(f _i)	(x _i)		
0-2	10	1	10	10
2-4	20	3	60	180
4-6	35	5	175	875
6-8	30	7	210	1470
8-10	5	9	45	405
	n = 100		Sum f _i x _i = 500	Sum f _i x _i ² = 2940

Solution

$$1) \bar{x} = \frac{\sum f_i x_i}{n} = \frac{500}{100} = 5.00$$

$$2) \text{s.d.}(\sigma_x) = \sqrt{\frac{\sum f_i x_i^2}{n} - (\bar{x})^2} = \sqrt{\frac{2940}{100} - (5.00)^2}$$

$$= \sqrt{29.40 - 25.00} = \sqrt{4.40} = 2.10 \text{ (\$in 100)}$$

Example Calculate s.d. of the marks of 100 students.

Marks	No. of students (f_i)	Mid-values (x_i)	$f_i x_i$	$f_i x_i^2$
0-2	10	1	10	10
2-4	20	3	60	180
4-6	35	5	175	875
6-8	30	7	210	1470
8-10	5	9	45	405
	$n = 100$		Sum $f_i x_i =$ 500	Sum $f_i x_i^2 =$ 2940

Solution

$$1) \quad \bar{x} = \frac{\sum f_i x_i}{n} = \frac{500}{100} = 5$$

$$\begin{aligned} \text{s.d. } (\sigma_x) &= \sqrt{\frac{\sum f_i x_i^2}{n} - (\bar{x})^2} \\ &= \sqrt{\frac{2940}{100} - (5)^2} = \sqrt{29.40 - 25} = \sqrt{4.40} \end{aligned}$$

$$2) \quad x \quad 2.21$$

Example The score of two teams A and B in 10 matches are as:

A	40	32	0	40	30	7	13	25	14	3
B	21	14	14	30	5	12	10	13	30	6

Find the variance for both the series. Which team is more consistent?

A			B		
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
40	19	361	21	4	16
32	11	121	14	-3	9
0	-21	441	14	12	144
40	19	361	30	13	169
30	9	81	5	-12	144
7	-8	196	12	-5	25
13	-8	64	10	-7	49
25	4	16	13	-4	16
14	-7	49	30	13	169
9	-12	144	6	-11	121
$\Sigma x_i = 210$		$\Sigma(x_i - \bar{x})^2 = 1834$	$\Sigma y_i = 110$		$\Sigma(y_i - \bar{y})^2 = 862$

Solution

$$1) \quad \bar{x} = \frac{\sum xi}{n} = \frac{210}{10} \quad \text{and} \quad \bar{y} = \frac{\sum yi}{n} = \frac{170}{10} = 17$$

$$\therefore \sigma_x = \sqrt{\frac{\sum (xi - \bar{x})^2}{n}} = \sqrt{\frac{1834}{10}} = \sqrt{183.4} = 13.54$$

$$\text{Also } \sigma_y = \sqrt{\frac{\sum (yi - \bar{y})^2}{n}} = \sqrt{\frac{862}{10}} = \sqrt{86.2} = 9.28$$

$$\text{Therefore, } (C.V.)_A = \frac{13.54}{21} \times 100 = 64.47\%$$

$$\text{and } (C.V.)_B = \frac{9.28}{17} \times 100 = 54.69\%$$

$$\text{Since } (C.V.)_A < (C.V.)_B$$

\therefore The base ball team A is more consistent.

5.4 Percentile

The nth percentile is that value (or size) such that n% of values of the whole data lies below it. For example, a score of 7% from the topmost score would be 93 the percentile as it is above 93% of the other scores.

Percentile Range

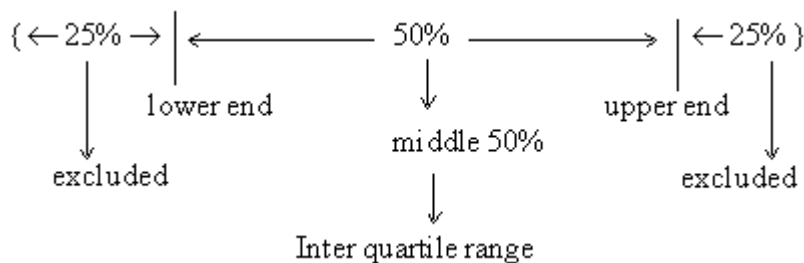
it is used as one of the measure of dispersion. it is a set of data and is defined as = P90 - P10 where P90 and P10 are the 90th and 10th percentile respectively. The semi - percentile range, i.e

$$\frac{P_{90} - P_{10}}{2}$$

can also be used but it is not common in use.

5.5 Quartiles and Interquartile Range

If we concentrate on two extreme values (as in the case of range), we don't get any idea about the scatter of the data within the range (i.e. the two extreme values). If we discard these two values the limited range thus available might be more informative. For this reason the concept of interquartile range is developed. It is the range which includes middle 50% of the distribution. Here 1/4 (one quarter of the lower end and 1/4 (one quarter) of the upper end of the observations are excluded.



Now the lower quartile (Q_1) is the 25th percentile and the upper quartile (Q_3) is the 75th percentile. It is interesting to note that the 50th percentile is the middle quartile (Q_2) which is in fact what you have studied under the title ' Median ". Thus symbolically

$$\text{Inter quartile range} = Q_3 - Q_1$$

If we divide ($Q_3 - Q_1$) by 2 we get what is known as Semi-inter quartile range.

i.e. $\frac{Q_3 - Q_1}{2}$. It is known as Quartile deviation (Q. D or SI QR).

$$\text{Therefore Q. D. (SI QR)} = \frac{Q_3 - Q_1}{2}$$

SAMPLE CAT

1. From the set of data given below,

3, 9, 5, 2, 7

- i) Find the mean and the median [2mks]
- ii) Calculate the standard deviation [3mks]
- iii) Calculate the geometric mean and the harmonic mean [4mks]

2. The following is the distribution of weights of 140 students of the Business statistics class of Mount Kenya University during the last intake.

Weight(in pounds)	Frequency
80 - 89	4
90 - 99	23
100 - 109	49
110 - 119	38
120 - 129	17
130 - 139	6
140 - 149	3

- i) Draw a histogram to represent the above information [4mks]
 - ii) Estimate the median weight in pounds [4mks]
 - iii) Calculate the interquartile range [4mks]
 - iv) Represent the information above by a box and whisker plot. [3mks]
3. Give a brief description of each of the sampling techniques listed below
- i) Random sampling [2mks]
 - ii) Stratified sampling [2mks]
 - iii) Systematic sampling. [2mks]

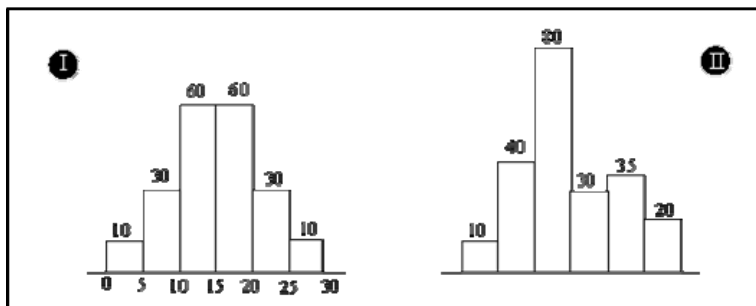
LECTURE 8

5.6 Skewness and Kurtosis

The voluminous raw data cannot be easily understood; hence, we calculate the measures of central tendencies and obtain a representative figure. From the measures of variability, we can know that whether most of the items of the data are close to our away from these central tendencies. But these statistical means and measures of variation are not enough to draw sufficient inferences about the data. Another aspect of the data is to know its symmetry. In the chapter "Graphic display" we have seen that a frequency may be symmetrical about mode or may not be. This symmetry is well studied by the knowledge of the "skewness." Still one more aspect of the curve that we need to know is its flatness or otherwise its top. This is understood by what is known as "Kurtosis."

Skewness

It may happen that two distributions have the same mean and standard deviations. For example, see the following diagram.



Although the two distributions have the same means and standard deviations they are not identical. Where do they differ?

They differ in symmetry. The left-hand side distribution is symmetrical one where as the distribution on the right-hand is asymmetrical or skewed. For a symmetrical distribution, the values, of equal distances on either side of the mode, have equal frequencies. Thus, the mode, median and mean - all coincide. Its curve rises slowly, reaches a maximum (peak) and falls

equally slowly (Fig. 1). But for a skewed distribution, the mean, mode and median do not coincide. Skewness is positive or negative as per the positions of the mean and median on the right or the left of the mode.

A positively skewed distribution (Fig.2) curve rises rapidly, reaches the maximum and falls slowly. In other words, the tail as well as median on the right-hand side. A negatively skewed distribution curve (Fig.3) rises slowly reaches its maximum and falls rapidly. In other words, the tail as well as the median are on the left-hand side.



5.6.1 Measure of Skewness

Pearson has suggested the use of this formula if it is not possible to determine the mode (Mo) of any distribution,

Then,

$$S_k = \frac{3(\text{mean} - \text{median})}{s.d}$$

Note : i) Although the co-efficient of skewness is always within $\square 1$, but Karl Pearson's co-efficient lies within ± 3 .

ii) $S_k = 0$, then there is no skewness

iii) If S_k is positive, the skewness is also positive.

iv) If S_k is negative, the skewness is also negative.

Unless and until no indication is given, you must use only Karl Pearson's formula.

Example The following table gives the frequency distribution of 291 workers of a factory according to their average monthly income in 1945 - 55. Calculate the median income.

Income group (\$)	No.of workers :
Below 50	1
50-70	16
70-90	39
90-110	58
110-130	60
130-150	46
150-170	22
170-190	15
190-210	15
210-230	9
230 & above	10

Solution:

Income group	f	c.f.
Below 50	1	1
50 – 70	16	17
70 – 90	39	56
90 – 110	58	114
110 – 130	60	174
130 – 150	46	220
150 – 170	22	242
170 – 190	15	257
190 – 210	15	252
210 – 230	9	281
230 & above	10	291
$n = \sum f = 291$		

Calculations :

1) Median = Size of $n + 1/2^{\text{th}}$ item

$$= \text{Size of } \frac{292}{2} \text{ item}$$

= Size of 146th item which lies in (100-130) class interval.

$$\text{Me} = \ell_1 + \left[\frac{n/2 - \text{c.f.}}{f} \right] (\ell_2 - \ell_1)$$

$$= 110 + \left[\frac{146 - 114}{60} \right] (130 - 100)$$

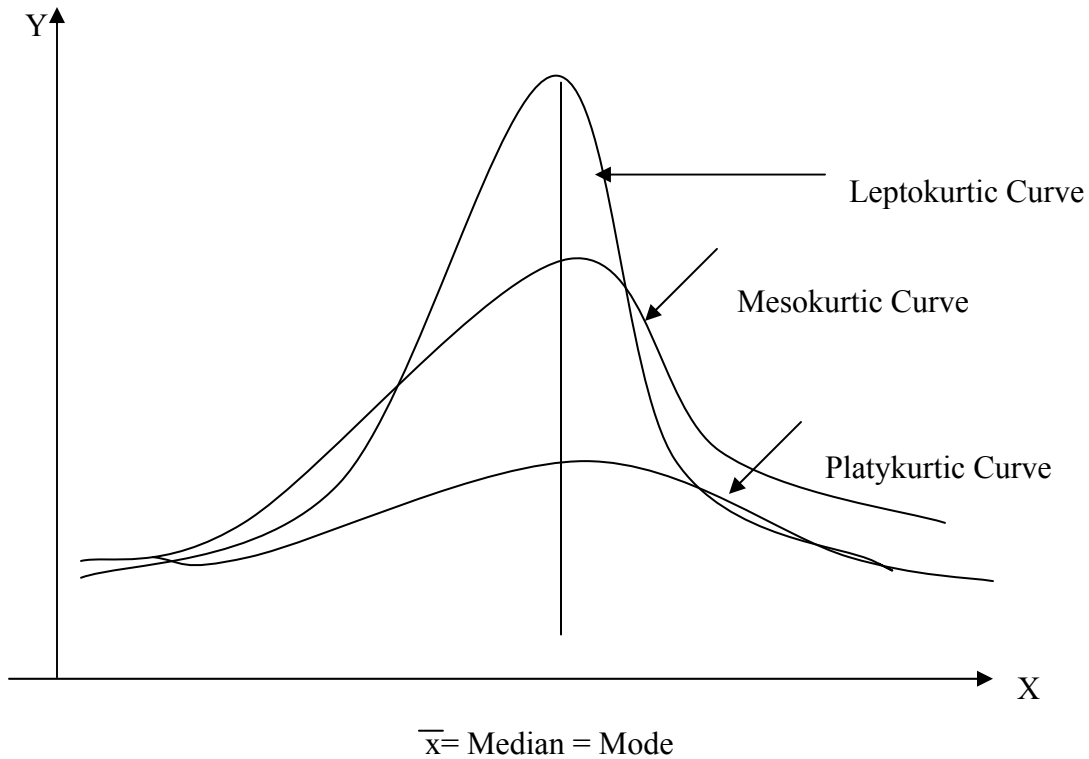
$$= 110 + \frac{31.5}{3}$$

$$= 120.5 \text{ dollars}$$

5.6.2 Kurtosis

It has its origin in the Greek word "Bulginess." In statistics it is the degree of flatness or 'peakedness' in the region of mode of a frequency curve. It is measured relative to the 'peakedness' of the normal curve. It tells us the extent to which a distribution is more peaked or flat-topped than the normal curve. If the curve is more peaked than a normal curve it is called 'Lepto Kurtic.' In this case items are more clustered about the mode. If the curve is more flat-topped than the more normal curve, it is Platy-Kurtic. The normal curve itself is known as "Meso Kurtic."

NATURE OF KURTOSIS



5.7 Characteristics of a good measure of Dispersion

- i. It should be easy to understand and to calculate.
- ii. It should be easy to interpret.
- iii. It should be open to further algebraic manipulation

Chapter Review Questions

1. The table shows the number of children in 50 families.

<i>Number of children</i>	<i>Frequency</i>	<i>Cumulative frequency</i>
<i>1</i>	<i>3</i>	<i>3</i>
<i>2</i>	<i>m</i>	<i>22</i>
<i>3</i>	<i>12</i>	<i>34</i>

<i>4</i>	<i>p</i>	<i>q</i>
<i>5</i>	<i>5</i>	<i>48</i>
<i>6</i>	<i>2</i>	<i>50</i>
	<i>T</i>	

- (a) Write down the value of *T*.
- (b) Find the values of *m*, *p* and *q*.

2. A group of 25 females were asked how many children they each had. The results are shown in the histogram below.

- (a) Show that the mean number of children per female is 1.4.

(2)

- (b) Show clearly that the standard deviation for this data is approximately 1.06.

(3)

- (c) Another group of 25 females was surveyed and it was found that the mean number of children per female was 2.4 and the standard deviation was 2. Use the results from parts (a) and (b) to describe the differences between the number of children the two groups of females have.

3. The following table shows the times, to the nearest minute, taken by 100 students to complete a mathematics task.

<i>Time (t) minutes</i>	<i>11–15</i>	<i>16–20</i>	<i>21–25</i>	<i>26–30</i>	<i>31–35</i>	<i>36–40</i>
<i>Number of students</i>	<i>7</i>	<i>13</i>	<i>25</i>	<i>28</i>	<i>20</i>	<i>7</i>

- (a) Construct a cumulative frequency table. (Use upper class boundaries 15.5, 20.5 and so on.)

- (b) On graph paper, draw a cumulative frequency graph, using a scale of 2 cm to represent 5 minutes on the horizontal axis and 1 cm to represent 10 students

on the vertical axis.

(c) Use your graph to estimate

(i) the number of students that completed the task in less than 17.5 minutes;

(ii) the time it will take for $\frac{3}{4}$ of the students to complete the task.

4. The table below shows the percentage, to the nearest whole number, scored by candidates in an examination.

<i>Marks (%)</i>	<i>0–9</i>	<i>10–19</i>	<i>20–29</i>	<i>30–39</i>	<i>40–49</i>	<i>50–59</i>	<i>60–69</i>	<i>70–79</i>	<i>80–89</i>	<i>90–100</i>
<i>Frequency</i>	<i>2</i>	<i>7</i>	<i>8</i>	<i>13</i>	<i>24</i>	<i>30</i>	<i>6</i>	<i>5</i>	<i>3</i>	<i>2</i>

The following is the cumulative frequency table for the marks.

<i>Marks (%)</i>	<i>Cumulative frequency</i>
<i>< 9.5</i>	<i>2</i>
<i>< 19.5</i>	<i>9</i>
<i>< 29.5</i>	<i>s</i>
<i>< 39.5</i>	<i>30</i>
<i>< 49.5</i>	<i>54</i>
<i>< 59.5</i>	<i>84</i>
<i>< 69.5</i>	<i>t</i>

< 79.5	95
< 89.5	98
< 100	100

(a) Calculate the values of s and of t .

(2)

(b) Using a scale of 1 cm to represent 10 marks on the horizontal axis, and 1 cm to represent 10 candidates on the vertical axis, draw a cumulative frequency graph.

(3)

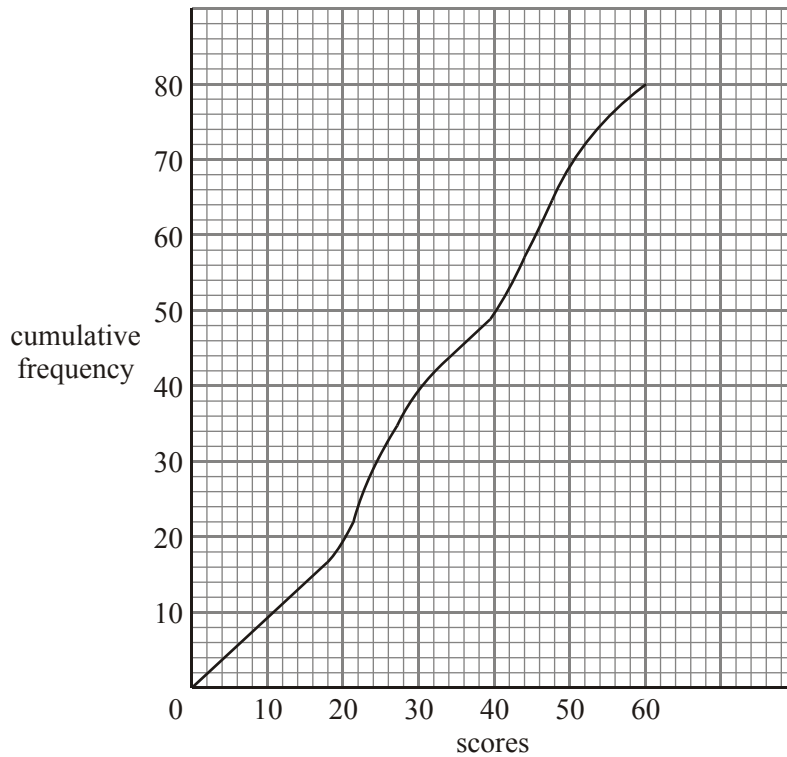
(c) Use your graph to estimate

(i) the median mark;

(ii) the lower quartile;

(iii) the pass mark, if 40% of the candidates passed.

5. The cumulative frequency graph below shows the examination scores of 80 students.



From the graph find

- (a) the median value;*
- (b) the interquartile range;*
- (c) the 35th percentile.*

6. The table below shows the number and weight (w) of fish delivered to a local fish market one morning.

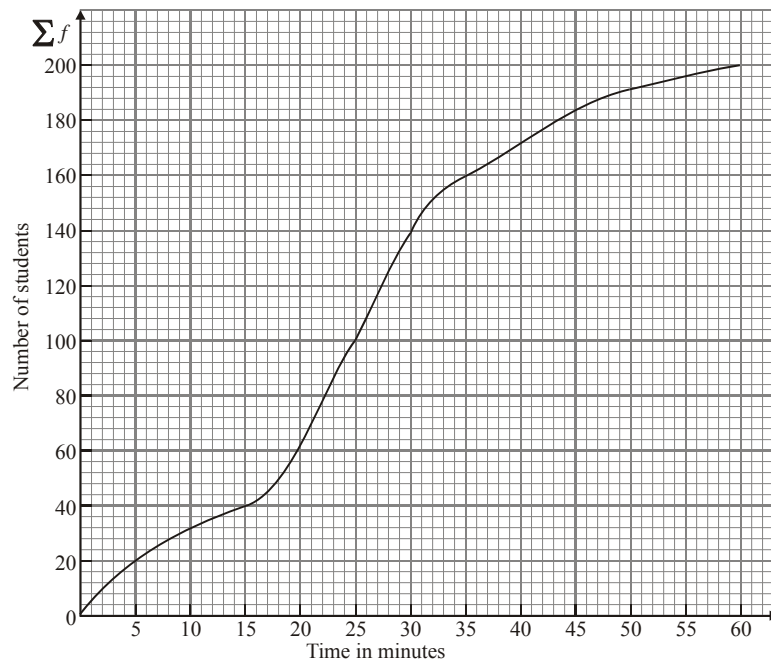
<i>weight (kg)</i>	<i>frequency</i>	<i>cumulative frequency</i>
$0.50 \leq w < 0.70$	<i>16</i>	<i>16</i>
$0.70 \leq w < 0.90$	<i>37</i>	<i>53</i>
$0.90 \leq w < 1.10$	<i>44</i>	<i>c</i>

$1.10 \leq w < 1.30$	23	120
$1.30 \leq w < 1.50$	10	130

- (a) (i) Write down the value of c . (1)
- (ii) On graph paper, draw the cumulative frequency curve for this data. Use a scale of 1 cm to represent 0.1 kg on the horizontal axis and 1 cm to represent 10 units on the vertical axis. Label the axes clearly. (4)
- (iii) Use the graph to show that the median weight of the fish is 0.95 kg. (1)
- (b) (i) The zoo buys all fish whose weights are above the 90th percentile. How many fish does the zoo buy? (2)
- (ii) A pet food company buys all the fish in the lowest quartile. What is the maximum weight of a fish bought by the company? (3)
- (c) A restaurant buys all fish whose weights are within 10% of the median weight.
- (i) Calculate the minimum and maximum weights for the fish bought by the restaurant. (2)
- (ii) Use your graph to determine how many fish will be bought by the restaurant

The cumulative frequency graph has been drawn from a frequency table showing the

time it takes a number of students to complete a computer game.



(a) From the graph find

- (i) the median time;**
- (ii) the interquartile range.**

The graph has been drawn from the data given in the table below.

<i>Time in minutes</i>	<i>Number of students</i>
$0 < x \leq 5$	20
$5 < x \leq 15$	20
$15 < x \leq 20$	<i>p</i>
$20 < x \leq 25$	40
$25 < x \leq 35$	60
$35 < x \leq 50$	<i>q</i>

$50 < x \leq 60$	10
------------------	----

- (b) *Using the graph, find the values of p and q .*
- (c) *Calculate an estimate of the mean time taken to finish the computer game.*

References

Business Calculations and statistics simplified by N.A Saleemi. Revised Edition. Pg 364-395

Essentials of statistics for Business and Economics by Anderson Sweety Williams Pg 69-81

LECTURE 9

CHAPTER 6: CORRELATION AND REGRESSION ANALYSIS

Purpose

To examine the relationship between two variables using correlation and to use regression analysis to interpolate or extrapolate.

Objectives

- a) Define correlation***
- b) Describe degrees of correlation***
- c) Identify methods of determining correlation***
- d) Compute Pearson's correlation coefficient.***
- e) Compute Spearman's rank correlation coefficient.***

6.1 Introduction to Correlation

So far we have considered only univariate distributions. By the averages, dispersion and skewness of distribution, we get a complete idea about the structure of the distribution. Many a time, we come across problems which involve two or more variables. If we carefully study the figures of rain fall and production of maize, figures of accidents and motor cars in a city, of demand and supply of a commodity, of sales and profit, we may find that there is some relationship between the two variables. On the other hand, if we compare the figures of rainfall in America and the production of cars in Japan, we may find that there is no relationship between the two variables. If there is any relation between two variables i.e. when one variable changes the other also changes in the same or in the opposite direction, we say that the two variables are correlated.

6.2 Correlation Defined

It means the study of existence, magnitude and direction of the relation between two or more variables.

6.3 Types of Correlation

1. Positive and negative correlation
2. Linear and non-linear correlation

A) If two variables change in the same direction (i.e. if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation. For example: Advertising and sales.

B) If two variables change in the opposite direction (i.e. if one increases, the other decreases and vice versa), then the correlation is called a negative correlation. For example: T.V. registrations and cinema attendance.

1. The nature of the graph gives us the idea of the linear type of correlation between two variables. If the graph is in a straight line, the correlation is called a "linear correlation" and if the graph is not in a straight line, the correlation is **non-linear** or **curvi-linear**.

For example, if variable x changes by a constant quantity, say 20 then y also changes by a constant quantity, say 4. The ratio between the two always remains the same (1/5 in this case). In case of a curvi-linear correlation this ratio does not remain constant.

6.4 Degrees of Correlation

Through the coefficient of correlation, we can measure the degree or extent of the correlation between two variables. On the basis of the coefficient of correlation we can also determine whether the correlation is positive or negative and also its degree or extent.

1. **Perfect correlation:** If two variables changes in the same direction and in the same proportion, the correlation between the two is **perfect positive**. According to Karl Pearson the coefficient of correlation in this case is +1. On the other hand if the variables change in the opposite direction and in the same proportion, the correlation is **perfect**

negative. its coefficient of correlation is -1. In practice we rarely come across these types of correlations.

2. **Absence of correlation:** If two series of two variables exhibit no relations between them or change in variable does not lead to a change in the other variable, then we can firmly say that there is **no correlation** or **absurd correlation** between the two variables. In such a case the coefficient of correlation is 0
3. **Limited degrees of correlation:** If two variables are not perfectly correlated or is there a perfect absence of correlation, then we term the correlation as Limited correlation. It may be positive, negative or zero but lies with the limits $\square 1$.

High degree, moderate degree or low degrees are the three categories of this kind of correlation. The following table reveals the effect (or degree) of coefficient or correlation.

Degrees	Positive	Negative
Absence of correlation	Zero	0
Perfect correlation	+ 1	-1
High degree	+ 0.75 to + 1	- 0.75 to -1
Moderate degree	+ 0.25 to + 0.75	- 0.25 to - 0.75
Low degree	0 to 0.25	0 to - 0.25

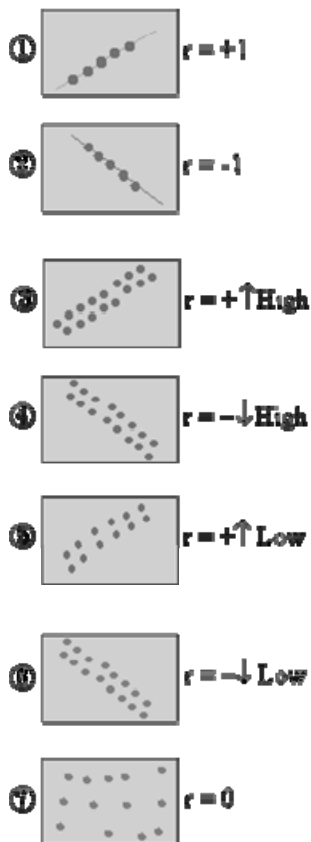
6.5 Methods Of Determining Correlation

We shall consider the following most commonly used methods

- (1) Scatter Plot
- (2) Karl Pearson's coefficient of correlation
- (3) Spearman's Rank-correlation coefficient.

6.5.1 Scatter Plot (Scatter diagram or dot diagram)

In this method the values of the two variables are plotted on a graph paper. One is taken along the horizontal (x-axis) and the other along the vertical (y-axis). By plotting the data, we get points (dots) on the graph which are generally scattered and hence the name 'Scatter Plot'. The manner in which these points are scattered, suggest the degree and the direction of correlation. The degree of correlation is denoted by 'r' and its direction is given by the signs positive and negative.



NOTES

i) If all points lie on a rising straight line the correlation is perfectly positive and $r = +1$ (see fig.1)

ii) If all points lie on a falling straight line the correlation is perfectly negative and $r = -1$ (see fig.2)

iii) If the points lie in narrow strip, rising upwards, the correlation is high degree of positive (see fig.3)

iv) If the points lie in a narrow strip, falling downwards, the correlation is high degree of negative (see fig.4)

v) If the points are spread widely over a broad strip, rising upwards, the correlation is low degree positive (see fig.5)

vi) If the points are spread widely over a broad strip, falling downward, the correlation is low degree negative (see fig.6)

vii) If the points are spread (scattered) without any specific pattern, the correlation is absent. i.e. $r = 0$. (see fig.7)

Though this method is simple and is a rough idea about the existence and the degree of correlation, it is not reliable. As it is not a mathematical method, it cannot measure the degree of correlation

6.5.2 Karl Pearson's coefficient of correlation

It gives the numerical expression for the measure of correlation. it is noted by ' r '. The value of ' r ' gives the magnitude of correlation and sign denotes its direction. Karl Pearson correlation coefficient is also sometimes referred to as the product moment correlation coefficient and is defined as

$$r = \frac{S_{xy}}{S_x S_y} \quad \text{Where}$$

$$s_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}, \quad s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}, \quad s_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}$$

s_{xy} is called the covariance of X and Y

s_x is the standard deviation of X

s_y is the standard deviation of Y

Therefore

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

Example 1

A chemical fertilizer company wishes to determine the extent of correlation between ‘quantity of compound X used’ and ‘lawn growth’ per day. The results are tabulated below:

	Compound	Lawn
Lawn	X (g)	Growth (mm)
A	1	3
B	2	3
C	4	6
D	5	8

Find the Pearson’s correlation between the three variables.

Solution

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

We start by obtaining the means \bar{x} and \bar{y} .

$$\bar{x} = \frac{1+2+4+5}{4} = \frac{12}{4} = 3$$

$$\bar{y} = \frac{3+3+6+8}{4} = \frac{20}{4} = 5$$

Now

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	3	-2	-2	4	4	4
2	3	-1	-2	2	1	4
4	6	1	1	1	1	1
5	8	2	3	6	4	9
				<u>13</u>	<u>10</u>	<u>18</u>

Substituting in the formula above,

$$r = \frac{13}{\sqrt{10}\sqrt{18}} = 0.969$$

A positive r means that as x (the mass of the chemical compound) increases, then so does y (the lawn growth)

A value of r close to 1 indicates a very strong positive correlation.

Alternative Formula for Calculating R

Often it is cumbersome to calculate the means when the data contains decimals or it is too large. A second formula that does not require the calculation of the means is shown below.

We have

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\text{COV}(xy)}{s_x s_y}$$

It can be shown that

$$s_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} \quad (\text{An alternative formula for finding the covariance of X and Y.})$$

$$s_x = \sqrt{\sum x^2 - \frac{(\sum x)^2}{n}} \quad (\text{An alternative formula for finding the standard deviation of X})$$

$$s_y = \sqrt{\sum y^2 - \frac{(\sum y)^2}{n}} \quad (\text{An alternative formula for finding the standard deviation of Y})$$

Therefore,

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)} \sqrt{\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

Using this formula, we can do example 1 above as follows:

$$x \quad y \quad (x)(y) \quad (x)^2 \quad (y)^2$$

1	3	12	1	9
2	3	6	4	9
4	6	24	16	36
5	8	40	25	64
12	20	73	46	118

Therefore

$$\begin{aligned}
 r &= \frac{73 - \frac{12 \times 20}{4}}{\left(46 - \frac{144}{4}\right) \left(118 - \frac{400}{4}\right)} \\
 &= \frac{13}{\sqrt{10} \sqrt{18}} \\
 &= 0.969
 \end{aligned}$$

Example 2

From the following data compute the coefficient of correlation between x and y.

- a) $s_x = 14.7$, $s_y = 19.2$, and $S_{xy} = 136.8$
- b) $\sum x = 65$, $\sum y = 141$, $\sum xy = 1165$, $\sum x^2 = 505$, $\sum y^2 = 2745$, $n = 11$

Solution

a)

$$r = \frac{s_{xy}}{s_x s_y} = \frac{136.8}{14.7 \times 19.2} = 0.485$$

b)

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)} \sqrt{\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$r = \frac{1165 - \frac{65 \times 141}{11}}{\sqrt{\left(505 - \frac{(65)^2}{11}\right)} \sqrt{\left(2745 - \frac{(141)^2}{11}\right)}}$$

$$\therefore r = 0.986$$

Example 3

If covariance between x and y is 12.3 and the variance of x and y are 16.4 and 13.8 respectively. Find the coefficient of correlation between them.

Solution: Given $\text{cov}(XY) = 12.3$, $s_x^2 = 16.4$ and $s_y^2 = 13.8$

$$r = \frac{\text{cov}(XY)}{s_x s_y}$$

$$r = \frac{12.3}{\sqrt{16.4}\sqrt{13.8}}$$

$$r = 0.818$$

6.5.3 Spearman's Rank Correlation Coefficient

This method is based on the ranks of the items rather than on their actual values. The advantage of this method over the others is that it can be used even when the actual values of items are unknown. For example if you want to know the correlation between honesty and wisdom of the boys of your class, you can use this method by giving ranks to the boys. It can also be used to find the degree of agreement between the judgments of two examiners or two judges. The formula is :

$$R = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

where R = Rank correlation coefficient

D = Difference between the ranks of two items

N = The number of observations.

Note: $-1 \leq R \leq 1$

- i) When R = +1: Perfect positive correlation or complete agreement in the same direction
- ii) When R = -1: Perfect negative correlation or complete agreement in the opposite direction.
- iii) When R = 0: No Correlation.

Computation:

- i. Give ranks to the values of items. Generally the item with the highest value is ranked 1 and then the others are given ranks 2, 3, 4, According to their values in the decreasing order.
- ii. Find the difference $D = R_1 - R_2$
where $R_1 = \text{Rank of } x$ and $R_2 = \text{Rank of } y$
Note that $\sum D = 0$ (always)
- iii. Calculate D^2 and then find $\sum D^2$
- iv. Apply the formula.

Note :

In some cases, there is a tie between two or more items. in such a case each items have ranks 4th and 5th respectively then they are given $\frac{4+5}{2} = 4.5$ th rank. If three items are of equal rank say 4th then they are given $\frac{4+5+6}{3} = 5$ th rank each. If m be the number of items of equal ranks, the factor $\frac{1}{12}[m^3 - m]$ is added to $\sum D^2$. If there are more than one of such cases then this factor added as many times as the number of such cases, then

$$R = 1 - \frac{\left\{ \sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \dots \right\}}{N(N^2 - 1)}$$

Example Calculate ‘ R ’ from the following data.

<u>Student</u> No.:	1	2	3	4	5	6	7	8	9	10
Rank in <u>Maths</u> :	1	3	7	5	4	6	2	10	9	8
Rank in Stats:	3	1	4	5	6	9	7	8	10	2

Solution :

Student No.	Rank in Maths (R_1)	Rank in Stats (R_2)	$R_1 - R_2$ D	$(R_1 - R_2)^2$ D^2
1	1	3	-2	4
2	3	1	2	4
3	7	4	3	9
4	5	5	0	0
5	4	6	-2	4
6	6	9	-3	9
7	2	7	-5	25

8	10	8	2	4
9	9	10	-1	1
10	8	2	6	36
N = 10			$\sum D = 0$	$\sum D^2 = 96$

Calculation of R :

$$R = 1 - \frac{6 \sum D^2}{N(N^2 - 1)} = 1 - \frac{6(96)}{10(100 - 1)}$$

$$= 1 - \frac{6 \times 96}{10 \times 99} = 0.4181$$

Example Calculate ' R ' of 6 students from the following data.

Marks in Stats :	40	42	45	35	36	39
Marks in English :	46	43	44	39	40	43

Solution:

Marks in Stats	R ₁	Marks in English	R ₂	R ₁ - R ₂	(R ₁ - R ₂) ² = D ²
40	3	46	1	2	4
42	2	43	3.5	-1.5	2.25
45	1	44	2	-1	1
35	6	39	6	0	0
36	5	40	5	0	0
39	4	43	3.5	0.5	0.25
N = 6				$\sum D = 0$	$\sum D^2 = 7.50$

Here m = 2 since in series of marks in English of items of values 43 repeated twice.

$\sum D^2$ is sometimes written as SD². [Read as sum of D squared]

$$R = 1 - \frac{6 \left\{ SD^2 + \frac{1}{12} (2^3 - 2) \right\}}{N(N^2 - 1)} = 1 - \frac{6 \left\{ 7.5 + \frac{1}{12} (8 - 2) \right\}}{6(36 - 6)}$$

$$R = 1 - \frac{6(7.5 + 0.5)}{210} = 0.771$$

Example The value of Spearman's rank correlation coefficient for a certain number of pairs of observations was found to be $2/3$. The sum of the squares of difference between the corresponding marks was 55. Find the number of pairs.

Solution: We have

$$R = 1 - \frac{6 SD^2}{N(N^2 - 1)} \quad \text{but } R = 2/3 \text{ and } SD^2 = 55$$

$$\therefore \frac{2}{3} = 1 - \frac{6 \times 55}{N(N^2 - 1)} \quad \therefore -\frac{1}{3} = -\frac{6 \times 55}{N(N^2 - 1)}$$

$$\therefore N(N^2 - 1) = 6 \times 55$$

$$\text{Now } N(N^2 - 1) = 990$$

$$\therefore N(N^2 - 1) = 10 \times 99 = 10(100 - 1)$$

$$\therefore N(N^2 - 1) = 10(10^2 - 1) \Rightarrow N = 10$$

Therefore, there were 10 students.

LECTURE 10

6.6 Introduction to Regression Analysis

Correlation gives us the idea of the measure of magnitude and direction between correlated variables. Now it is natural to think of a method that helps us in estimating the value of one variable when the other is known. Also correlation does not imply causation. The fact that the variables x and y are correlated does not necessarily mean that x causes y or vice versa. For example, you would find that the number of schools in a town is correlated to the number of accidents in the town. The reason for these accidents is not the school attendance; but these two increases what is known as **population**. A statistical procedure called **regression** is concerned with causation in a relationship among variables. It assesses the contribution of one or more variable called **causing variable** or **independent variable** or one which is being **caused** (**dependent variable**). When there is only one independent variable then the relationship is expressed by a straight line. This procedure is called **simple linear regression**.

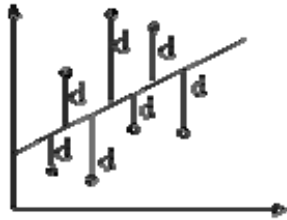
Regression can be defined as a method that estimates the value of one variable when that of other variable is known, provided the variables are correlated. The dictionary meaning of regression is "to go backward." It was used for the first time by Sir Francis Galton in his research paper "Regression towards mediocrity in hereditary stature."

6.7 Lines of Regression

In scatter plot, we have seen that if the variables are highly correlated then the points (dots) lie in a narrow strip. if the strip is nearly straight, we can draw a straight line, such that all points are close to it from both sides. such a line can be taken as an ideal representation of variation. This line is called the line of best fit if it minimizes the distances of all data points from it.

This line is called **the line of regression** or line of best fit. Now prediction is easy because now all we need to do is to extend the line and read the value. Thus to obtain a line of regression, we need to have a line of best fit. But statisticians don't measure the distances by dropping perpendiculars from points on to the line. They measure **deviations** (or **errors** or **residuals** as they are called) (i) vertically and (ii) horizontally.

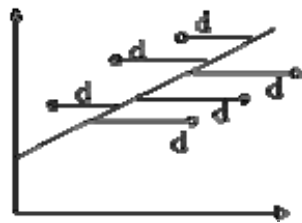
Thus we get two lines of regressions as shown in the figure (1) and (2).



(1) Line of regression of y on x

Its form is $y = a + b x$

It is used to estimate y when x is given



(2) Line of regression of x on y

Its form is $x = a + b y$

It is used to estimate x when y is given.

6.8 Regression Equation of y on x

It can be obtained by (1) graphically - by Scatter plot (ii) Mathematically - by the method of least squares.

The least squares formula is

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x})$$

Example 1

Use the least squares formula to fit a regression line through (1,3), (3,5) and (5,6).

Solution

x	y	$(x)(y)$	$(x)^2$
1	3	3	1
3	5	15	9
5	6	30	25
9	14	48	35

So

$$\sum x = 9, \sum y = 14, \sum xy = 48, \sum x^2 = 35, n = 3$$

$$\therefore \bar{x} = \frac{\sum x}{n} = \frac{9}{3} = 3 \text{ and } \bar{y} = \frac{\sum y}{n} = \frac{14}{3}$$

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x})$$

$$s_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$s_{xy} = 48 - \frac{9 \times 14}{3} = 6$$

$$s_x^2 = \sum x^2 - \frac{(\sum x)^2}{n} = 35 - \frac{9^2}{3} = 8$$

Therefore,

$$y - \frac{14}{3} = \frac{6}{8} (x - 3)$$

Multiplying through by 24 (L.C.M),

$$24y - 112 = 18x - 54$$

$$24y = 18x - 54 + 112$$

$$24y = 18x + 58$$

$$y = \frac{18}{24}x + \frac{58}{24}$$

$$y = 0.75x + 2.42$$

Example 2

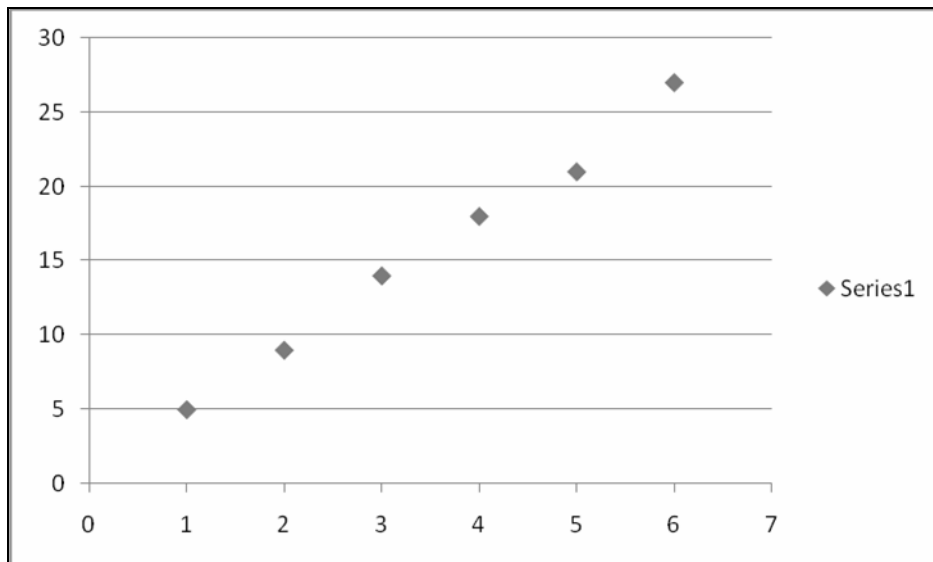
The table below shows the sales for Bidii electronics established in the late 1998.

Year	1999	2000	2001	2002	2003	2004
Sales(Sh x1000)	5	9	14	18	21	27

- Draw a scatter graph to represent this data.
- Find r^2
- Find the equation of the line of best fit using the linear regression formula.
- Predict the sales for the year 2006, giving your answer to the nearest Sh

Solution

a)



b) To get r^2

x	y	$(x)(y)$	$(x)^2$	$(y)^2$
1	5	5	1	25
2	9	18	4	81
3	14	42	9	196
4	18	72	16	324
5	21	105	25	441
6	27	162	36	729
21	94	404	91	1796

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)} \sqrt{\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$r = \frac{404 - \frac{21 \times 94}{6}}{\sqrt{\left(91 - \frac{(21)^2}{6}\right)} \sqrt{\left(1796 - \frac{94^2}{6}\right)}}$$

$$= \frac{404 - 329}{\sqrt{(91 - 73.5)} \sqrt{(1796 - 1472.66)}} = \frac{75}{75.22}$$

$$r = 0.997$$

$$\therefore r^2 = 0.994$$

c)

$$\bar{x} = \frac{21}{6} = 3.5 \quad , \quad \bar{y} = \frac{94}{6} = 15.67$$

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x})$$

$$s_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$s_{xy} = 404 - \frac{21 \times 94}{6} = 75$$

$$s_x^2 = \sum x^2 - \frac{(\sum x)^2}{n} = 91 - \frac{21^2}{6} = 17.5$$

$$y - 15.67 = \frac{75}{17.5} (x - 3.5)$$

$$y = \frac{75}{17.5} x - 15 + 15.67$$

Thus is the equation for the regression line of y on x.

$$y = 4.29x + 0.67$$

In the year 2008, $x = 8$

Therefore

$$y = (4.29)(8) + 0.67$$

$$y = 34.32 + 0.67$$

$$\text{Sales} = 34.99 \times 1,000$$

$$34,990$$

= Sh 35000 worth of sales.

Example3

A panel of two judges A and B graded dramatic performance by independently awarding marks as follows:

Performance No.:	1	2	3	4	5	6	7	8
Marks by A	: 36	32	34	31	32	32	35	38
Marks by B	: 35	33	31	30	34	32	36	?

- Obtain the correlation coefficient r
- Use the least squares method to obtain the regression equation of y on x .
- Find the mark awarded by judge B to performance 8.

Solution:

a)

x	y	xy	x^2	y^2
36	35	1260	1296	1225
32	33	1056	1024	1089
34	31	1054	1156	961
31	30	930	961	900
32	34	1088	1024	1156
32	32	1024	1024	1024

35	36	1260	1225	1296
232	231	7672	7710	7651

$$\therefore \bar{x} = \frac{\sum x}{n} = \frac{232}{7} = 33 \text{ and } \bar{y} = \frac{\sum y}{n} = \frac{231}{7} = 33$$

Now

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)} \sqrt{\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$r = \frac{7672 - \frac{232 \times 231}{7}}{\sqrt{\left(7710 - \frac{(232)^2}{7}\right)} \sqrt{\left(7651 - \frac{231^2}{7}\right)}}$$

$$\frac{7612 - 7656}{\sqrt{(7710 - 7689)} \sqrt{(7651 - 7623)}} = \frac{16}{\sqrt{21} \times \sqrt{28}}$$

Therefore $r=0.65$

b) The equation of the line of regression of y on x

$$y - \bar{y} = \frac{s_{xy}}{s_x} (x - \bar{x})$$

$$s_{xy} = 16 \quad s_x^2 = 21$$

$$y - 33 = \frac{16}{21} (x - 33)$$

$$\Rightarrow y = 0.76x + 7.92$$

c) Inserting $x = 38$, we get

$$y = 0.76 (38) + 7.92$$

$$y = 36.8 = 37 (\text{approximately})$$

Therefore, the Judge B would have given 37 marks to 8th performance

Alternative Formula for Calculating Regression

It is expressed as $y = a + bx$ where a and b are two unknown constants which determine the position of the line completely.

If the values of a and b are completely determined the equation of the regression line of y on x is obtained.

The two basic equations which can be solved simultaneously to find a and b are:-

$$\sum y = na + b \sum x \dots\dots\dots(i)$$

$$\sum xy = a \sum x + b \sum x^2 \dots\dots(ii)$$

Example 4

From 10 observations of price x and supply y of a commodity the results obtained

$$\sum x = 130, \sum y = 220, \sum x^2 = 2288, \sum xy = 3467$$

Compute the regression of y on x and interpret the result. Estimate the supply when the price of 16 units.

Solution: The equation of the line of regression of y on x

$$y = a + b x$$

Also from normal equations

$$\sum y = n a + b \sum x \text{ and } \sum xy = a \sum x + b \sum x^2$$

we get

$$220 = 10 a + 130 b \dots \dots \dots (1)$$

$$3467 = 130 a + 2288 \dots \dots \dots (2)$$

Solving (1) and (2) a

$$2860 = 130 a + 1690 b$$

$$3467 = 130 a + 2288 b$$

On subtraction,

$$607 = 598 b$$

$$b = 1.002$$

Putting $b = 1.002$ in $220 = 10 a + 130 b$, we get $a = 8.974$.

Hence the 3 equation of the line of regression of y on x is

$$y = 8.974 + 1.002 x$$

When $x = 16$, we get

$$y = 8.974 + 1.002 (16)$$

$$y = 25.006$$

6.9 Uses of Regression Analysis

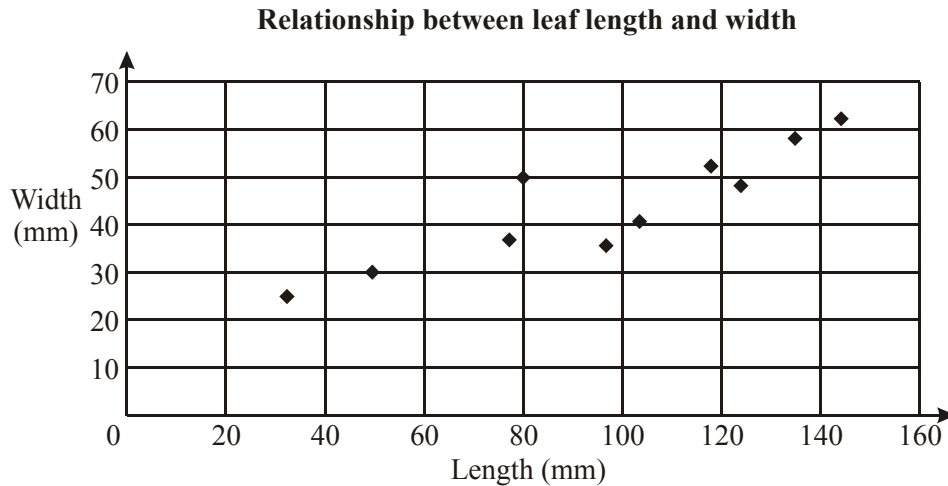
1. Through the methods of interpolation and extrapolation, it provides estimates of values of the dependent variable from the values of the independent variables.
2. Regression analysis also enables us to obtain a measure of error involved in using the regression line as a basis of estimation.
3. Regression analysis also enables us to compute the coefficient of determination- which gives the measure of association or correlation between two variables.

6.10 Difference between correlation and regression

- i) The objective of regression analysis is to study the relationship between the variables involved while the coefficient of correlation is the measure of the degree of relationship between the variables.
- ii) The cause and effect relation is clearly indicated through regression analysis but we cannot say that one variable is the cause and the other the effect.
- iii) Correlation analysis is only confined to the study of linear relationship between the variables, and therefore has limited applications. Regression analysis has much wider applications as it studies both linear and nonlinear relationship between variables.
- iv) There may be nonsense correlation between two variables which is due to mere chance and has no practical relevance, but there is no such thing as nonsense regression.

Chapter Review Questions

1. The length and width of 10 leaves are shown on the scatter diagram below.



(a) Draw a suitable line of best fit.

(b) Write a sentence describing the relationship between leaf length and leaf width for this sample.

2. Statements I, II, III, IV and V represent descriptions of the correlation between two variables.

I High positive linear correlation

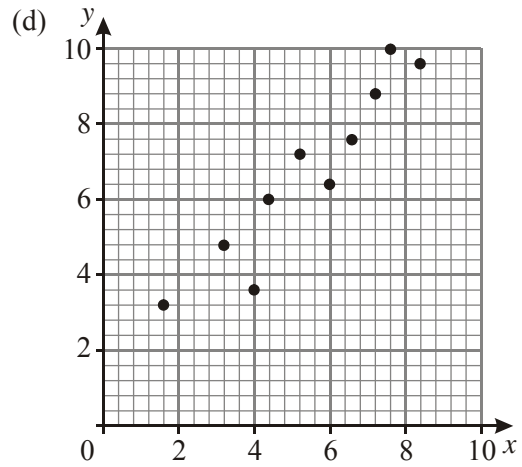
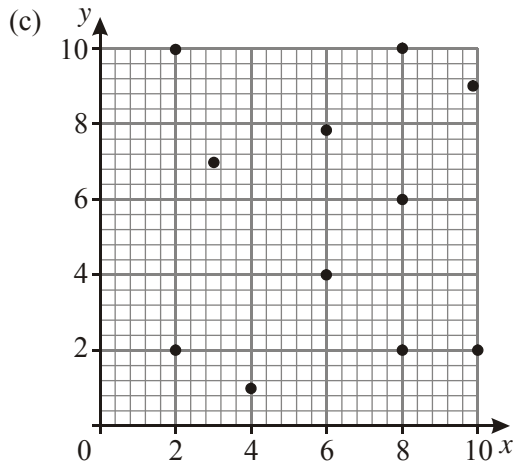
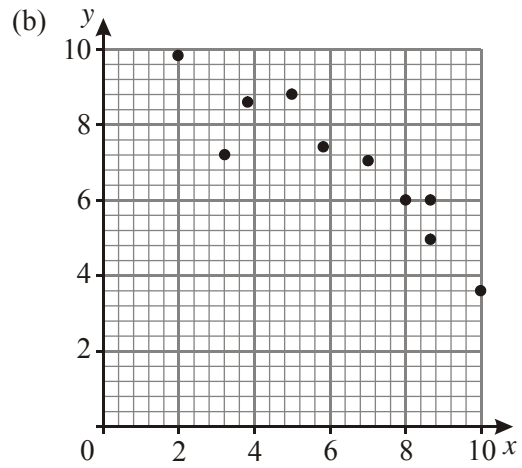
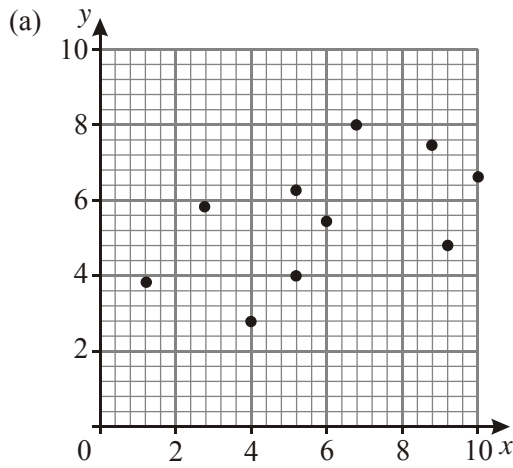
II Low positive linear correlation

III No correlation

IV Low negative linear correlation

V High negative linear correlation

Which statement best represents the relationship between the two variables shown in each of the scatter diagrams below.



Answers:

(a)

(b)

(c)

(d)

2. The Type Fast secretarial training agency has a new computer software spreadsheet package. The agency investigates the number of hours it takes people of varying ages to reach a level of proficiency using this package. Fifteen

individuals are tested and the results are summarised in the table below.

<i>Age</i> (<i>x</i>)	32	40	21	45	24	19	17	21	27	54	33	37	23	45	18
<i>Time</i> (in hours) (<i>y</i>)	10	12	8	15	7	8	6	9	11	16	<i>t</i>	13	9	17	5

(a) (i) Given that $S_y = 3.5$ and $S_{xy} = 36.7$, calculate the product-moment correlation coefficient r for this data.

(ii) What does the value of the correlation coefficient suggest about the relationship between the two variables?

(b) Given that the mean time taken was 10.6 hours, write the equation of the regression line for y on x in the form $y = ax + b$.

(c) Use your equation for the regression line to predict

(i) the time that it would take a 33 year old person to reach proficiency, giving your answer correct to the nearest hour;

(ii) the age of a person who would take 8 hours to reach proficiency, giving your answer correct to the nearest year.

3. Ten students were given two tests, one on Mathematics and one on English.

The table shows the results of the tests for each of the ten students.

<i>Student</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>
<i>Mathematics</i> <i>(x)</i>	8.6	13.4	12.8	9.3	1.3	9.4	13.1	4.9	13.5	9.6
<i>English</i> <i>(y)</i>	33	51	30	48	12	23	46	18	36	50

(a) Given s_{xy} (the covariance) is 35.85, calculate, correct to two decimal places, the product moment correlation coefficient (r).

(6)

(b) Use your result from part (a) to comment on the statement:

'Those who do well in Mathematics also do well in English.'

4. The heights and weights of 10 students selected at random are shown in the table below.

<i>Student</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
<i>Height</i> <i>x cm</i>	155	161	173	150	182	165	170	185	175	145
<i>Weight</i> <i>y kg</i>	50	75	80	46	81	79	64	92	74	108

(a) Plot this information on a scatter graph. Use a scale of 1 cm to represent 20 cm on the x -axis and 1 cm to represent 10 kg on the y -axis.

(b) Calculate the mean height.

(c) Calculate the mean weight.

(d) It is given that $S_{xy} = 44.31$.

(i) By first calculating the standard deviation of the heights, correct to two decimal places, show that the gradient of the line of regression of y on x is 0.276.

(ii) Calculate the equation of the line of best fit.

(iii) Draw the line of best fit on your graph.

(e) Use your line to estimate

(i) the weight of a student of height 190 cm;

(ii) the height of a student of weight 72 kg.

(f) It is decided to remove the data for student number 10 from all calculations. Explain briefly what effect this will have on the line of best fit.

References

- i. *Research methods* by Mugenda Olive M and Mugenda Abel G. Pg132-134**
- ii. *Business Calculations and statistics simplified* by N.A Saleemi. Revised Edition. Pg 480-501, 508-522**
- iii. *Essentials of statistics for Business and Economics* by Anderson Sweety Williams Pg 84-9**

LECTURE 11

CHAPTER 7: RANDOM VARIABLES AND THEIR PROBABILITY DISTRIBUTIONS



Purpose

To acquaint the student with knowledge of various probability distributions and to use them to solve various problems related to statistics.

Objectives

- a) *Define a random variable.*
- b) *State and describe the features of the following distributions:*
 - i. *Binomial distribution*
 - ii. *Poisson distribution*
 - iii. *Normal distribution*
- c) *Use tables to read probabilities for the above distributions.*

7.1 Introduction

Any variable can have a number of possible values. If the values occur unpredictably, it may be called a **random variable**. Each value (in case of a discrete variable) may have a particular chance or probability of occurring. The sum of the probabilities of all the possible values is always 1. Recall ‘discrete’ means ‘positive whole number’. The variable we are dealing with can assume a discrete value. Example: We may have zero, one, two.....people in a room. The number would be limited by the size of the room. We cannot have negative two people or one and a half people.

Example : The number of people (X) in a moving car licensed to carry up to 6 people follows the pattern below: -

Number Probability

X	P (X)
1	0.5
2	0.3

3	0.1
4	0.05
5	0.03
6	<u>0.02</u>

1.0 Total probability

From this 'distribution' we can find any of the following probabilities:

Exactly 3 people: $P(X = 3) = 0.1$

Up to 3 people: $P(X \leq 3) = 0.5 + 0.3 + 0.1 = 0.9$

Fewer than 3 people: $P(X < 3) = 0.5 + 0.3 = 0.8$

More than 3 people: $P(X > 3) = 0.05 + 0.03 + 0.02 = 0.1$

At least 3 people: $P(X \geq 3) = 0.1 + 0.05 + 0.03 + 0.02 = 0.2$

Between 2 and 5 people: $P(2 \leq X \leq 5)^* = .3 + .1 + .05 + .03 = .48$

Anything but 3 people: $P(X \neq 3) = .5 + .3 + .05 + .05 + .03 + .02 = .9$

* $P(2 \leq X \leq 5)$ is read as X is greater than or equal to 2 but less than or equal to 5.

7.2 The Binomial Distribution

The binomial distribution is used when there are exactly two mutually exclusive outcomes of a trial. These outcomes are appropriately labeled "success" and "failure". The binomial distribution is used to obtain the probability of observing x successes in N trials, with the probability of success on a single trial denoted by p. The binomial distribution assumes that p is fixed for all trials.

7.2.1 Features of a Binomial Distribution

The **Binomial** experiment consists of:

- A fixed number of trials, n
- Two possible outcomes, p and q (or $1 - p$). p is called ‘success’ and $1 - p$ (the complement) is called ‘failure’. ‘Success’ is what we are interested in. For example, the proportion of defective items produced by a factory in a day may be 1%. $p = 0.01$.
- Independent trials (the outcome of one trial does not affect the outcomes of any other trials).

7.2.2 Using the Binomial Tables:

Different probabilities can be read from the binomial tables.

Each page focuses on a given value of ‘ n ’. Across the top are various values of ‘ p ’. In the left margin are the possible values of ‘ k ’. The body of the table gives **$P(X \leq k)$** .

Example 1

Find the probability of at least two defective items in a batch of 10 items with a defective rate of 10%.

Here, we would need to find the following:-

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

If we used $P(X \leq k)$ we would only need to find:

$$P(X \geq 2) = 1 - P(X \leq 1)$$

Your binomial tables are at the back of this Module.

Example 2:

A coin is tossed 10 times. What is the probability that there will be not more than 4 'heads'?

Solution: $n=10$, $p=.5$ and we require $P(X \leq 4)$

Directly from the tables using $k=4$ we get $P(X \leq 4) = 0.377$

Example 3: 15 children are born. What is the probability that there will be more than 9 girls?

Solution: $n=15$, $p=.5$ and we require $P(X > 9)$

$$P(X > 9) = P(X \geq 10) = 1 - P(X \leq 9)$$

$$= 1 - 0.849$$

$$= 0.151 \text{ (this may be written as 0.151 or .151).}$$

Example 3: The chance of any computer chip being defective is 20%. If 15 chips are selected at random, what is the probability that

- a) Fewer than 3 will be defective?
- b) None will be defective?
- c) Exactly 5 will be defective?
- d) Between 6 and 8 will be defective?
- e) At least 10 will not be defective?

Solutions: In all cases $n = 15$, $p = 0.2$

Reading from the Binomial tables,

$$\begin{aligned} \text{a) } P(X < 3) &= P(X \leq 2) \\ &= 0.398 \end{aligned}$$

$$\begin{aligned} \text{b) } P(X = 0) &= P(X \leq 0) \\ &= 0.035 \end{aligned}$$

$$\begin{aligned}
 \text{c) } P(X = 5) &= P(X \leq 5) - P(X \leq 4) \\
 &= 0.939 - 0.836 \\
 &= 0.103
 \end{aligned}$$

$$\begin{aligned}
 \text{d) } P(6 \leq X \leq 8) &= P(X \leq 8) - P(X \leq 5) \\
 &= 0.999 - 0.939 \\
 &= 0.06
 \end{aligned}$$

e) 'At least 10 not defective' = 'at most 5 defective'

$$P(X \leq 5) = 0.939$$

Alternatively, we could have used $n=15$, $p=0.8$ (not defective)*

$$\begin{aligned}
 P(X \geq 10) &= 1 - P(X \leq 9) \\
 &= 1 - 0.061 \\
 &= 0.939 \text{ which is the same result.}
 \end{aligned}$$

Note: *The designations for success and failure can be interchanged

So far we have shown how we determine the probabilities of a binomial experiment by using the Binomial Tables. What if the number of trials, n , or the probability, p , cannot be found in the tables? In this situation we can use formula or Excel (insert – function. Select 'statistical' and 'binomdist'). However, you are not required to use formula in this course. You will use the tables. For your *information* only the formula is:

$$P = (x, p, n) = \binom{n}{x} (p)^x (1-p)^{(n-x)} \quad \text{for } 0, 1, 2, \dots, n$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

X represents the name of the random variable. x represents the value of the random variable. The factorial sign (!) is best explained by example. 6! means 6 x 5 x 4 x 3 x 2 x 1 = 720. You will find the ! button on your calculator.

7.3 The Poisson Distribution

A particular event occurs ‘on average’ μ times within a certain period or situation.

What is the probability that it will occur k times in that period?

A Poisson probability distribution is the number of occurrences per interval of **time** or **space**.

Time? Example: A programmer makes, on average, 2 mistakes per hour. Space?

Example: A programmer makes, on average, 2 mistakes per program.

Example : The average number of arrivals at a doctor’s clinic is 3 per hour. What is the probability that in a given hour there will be 5 arrivals? Here we have $\mu=3$ and $k=5$.

Note:

1) There is no maximum possible value for ‘k’.

7.3.1 Features of the Poisson Experiment

- The number of successes that occur in a period of time or an interval of space is independent of the number of successes that occur in any other interval

- The probability of a success in an interval is the same for all *equal-sized* intervals and proportional to the size of the interval (for example an average of 4 in an hour is equivalent to an average of 2 in 30 minutes (1/2 hour) and 8 in two hours).

Example: The possible number of accidents on a given stretch of road is infinite (only restricted by time).

2) The stated mean applies only to the given specific situation or time period. If the time period or the size of the situation changes, the mean must be adjusted accordingly.

Example: If 4 trucks cross a bridge in 20 minutes (on average), then 2 trucks will cross the bridge in a 10 minute period (on average).

7.3.2 Using the Poisson Tables:

Across the top are various values of ‘ μ ’. In the left margin are the possible values of ‘ k ’. The body of the table gives $P(X \leq k)$.

Eg. If $\mu = 3$ and $k = 5$ then $P(X \leq 5) = .916$.

Note: This is not the probability of ‘exactly 5’ arrivals, but ‘up to 5’ arrivals.

Illustration : Poisson Tables for μ values are at the back of this module.

Example 1

The average number of faults in each TV set is 3. What is the probability that a TV set chosen at random will have a) not more than 4 faults? b) more than 6 faults? c) between 3 and 5 faults? d) exactly 2 faults?

Solutions : For all cases $\mu = 3$.

a) $P(X \leq 4) = 0.815$

$$\begin{aligned}
 \text{b) } P(X > 6) &= P(X \geq 7) = 1 - P(X \leq 6) \\
 &= 1 - 0.966 \\
 &= 0.034
 \end{aligned}$$

$$\begin{aligned}
 \text{c) } P(3 \leq X \leq 5) &= P(X \leq 5) - P(X \leq 2) \\
 &= 0.916 - 0.423 \\
 &= 0.493
 \end{aligned}$$

$$\begin{aligned}
 \text{d) } P(X = 2) &= P(X \leq 2) - P(X \leq 1) \\
 &= 0.423 - 0.199 \\
 &= 0.244
 \end{aligned}$$

Example 2

The average number of telephone calls you receive in your office every hour is 3.

a) You must leave the office on urgent business that is expected to take half an hour. What is the probability that you will not miss any calls during that time?

b) In fact the 'urgent business' takes two hours. What is the probability that you have missed at least 5 calls?

Solutions: Average number of calls per hour is 3.

a) $\mu = 1.5$ (for half an hour)

$$\begin{aligned}
 P(X = 0) &= P(X \leq 0) = P(X = 0) \text{ because there are no values } < 0 \\
 &= 0.223
 \end{aligned}$$

b) $\mu = 6.0$ (for two hours)

$$\begin{aligned}
 P(X \geq 5) &= 1 - P(X \leq 4) \\
 &= 1 - 0.285 \\
 &= 0.715
 \end{aligned}$$

Example 3

The average number of fire-crews engaged in fighting fires in a large city at any one time is 2.5. How many fire-crews should be available on duty so that they can respond to at least 95% of all fire emergencies?

Solution :

Given $\mu = 2.5$, we need to find a value of k such that $P(X \leq k) > .95$

K	P(X ≤ k)
3	.758
4	.891
5	.958

Five crews are needed to cover at least 95% of emergencies.

This latter problem is a ‘poisson-reverse’ question. Normally the question *asks* for a probability and *gives* the x or k value. Here the probability is *given* and the question requires k.

So far we have shown how we determine the probabilities of a Poisson experiment

By using the Poisson Tables. What if the value of the population mean (μ) cannot be found in the tables? In this situation we can use formula or Excel (insert – function)

Select ‘statistical’ and ‘Poisson’). **However, you are not required to use formula in this course. You will use the tables.** For your *information* only the formula is:

$$X = \frac{\lambda^x e^{-\lambda}}{x!}$$

λ is the mean and is sometimes represented by μ

e is the base of the natural logarithm (approximately 2.71828).

7.3.3 Conditions under which a Poisson distribution holds

- counts of rare events
- all events are independent
- average rate does not change over the period of interest

7.3.4 Examples of experiments where a Poisson distribution holds:

- birth defects
- number of sample defects on a car
- number of typographical errors on a page
- **Examples of Poisson probability distribution:**
- The mass of alpha particles released by a radioactive source in a known interval of time.
- The number of phone calls approved at a telephone exchange in a known interval of time.
- The amount of imperfect research paper in a packet of 100, created by a good industry.
- The amount of printing errors at every page of a book by a good publication.
- The number of road accidents reports in a city at a particular junction at a particular tim

7.3.5 Examples of experiments where a Poisson distribution may not hold

- number of insects on a tree - contagion?
- number of males in families of size 4 - not 'rare' events

LECTURE 12

7.4 The Normal Distribution

Normal distribution is probably one of the most important and widely used continuous distribution. It is known as a normal random variable, and its probability distribution is called a normal distribution. The following are the characteristics of the normal distribution:

7.4.1 Features of the Normal Distribution:

1. It is bell shaped and is symmetrical about its mean.
2. It is asymptotic to the axis, i.e., it extends indefinitely in either direction from the mean.
3. It is a continuous distribution.
4. It is a family of curves, i.e., every unique pair of mean and standard deviation defines a different normal distribution. Thus, the normal distribution is completely described by two parameters: mean and standard deviation. See the following figure.
5. Total area under the curve sums to 1, i.e., the area of the distribution on each side of the mean is 0.5.
6. It is unimodal, i.e., values mound up only in the center of the curve.
7. The probability that a random variable will have a value between any two points is equal to the area under the curve between those points.

7.5 The Standard Normal Distribution

Note that the integral calculus is used to find the area under the normal distribution curve. However, this can be avoided by transforming all normal distribution to fit the standard normal distribution. This conversion is done by rescaling the normal distribution axis from its true units (time, weight, dollars, and...) to a standard measure called Z score or Z value. A Z score is the number of standard deviations that a value, X, is away from the mean. If the value of X is greater than the mean, the Z score is positive; if the value of X is less than the mean, the Z score is negative. The Z score or equation is as follows:

$$Z = (X - \text{Mean}) / \text{Standard deviation}$$

A standard Z table can be used to find probabilities for any normal curve problem that has been converted to Z scores. The Z distribution is a normal distribution with a mean of 0 and a standard deviation of 1.

- Probability Function given by

Standard Normal	Non-Standard Normal
<p style="text-align: center;">Mean = 0 <i>and</i> Variance = 1</p>	<p style="text-align: center;">Mean is not 0 <i>or</i> Variance is not 1</p>

The following steps are helpful when working with the normal curve problems:

1. Graph the normal distribution, and shade the area related to the probability you want to find.
2. Convert the boundaries of the shaded area from X values to the standard normal random variable Z values using the Z formula above.
3. Use the standard Z table to find the probabilities or the areas related to the Z values in step 2.

Example 1

Graduate Management Aptitude Test (GMAT) scores are widely used by graduate schools of business as an entrance requirement. Suppose that in one particular year, the mean score for the GMAT was 476, with a standard deviation of 107. Assuming that the GMAT scores are normally distributed, answer the following questions:

Question 1. What is the probability that a randomly selected score from this GMAT falls between 476 and 650? $476 \leq x \leq 650$ the following figure shows a graphic representation of this problem.

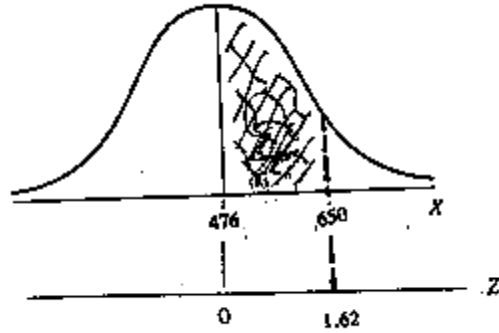


Figure 4

Applying the Z equation, we get: $Z = (650 - 476)/107 = 1.62$. The Z value of 1.62 indicates that the GMAT score of 650 is 1.62 standard deviation above the mean. The standard normal table gives the probability of value falling between 650 and the mean. The whole number and tenths place portion of the Z score appear in the first column of the table. Across the top of the table are the values of the hundredths place portion of the Z score. Thus the answer is that 0.4474 or 44.74% of the scores on the GMAT fall between a score of 650 and 476.

Question 2.

What is the probability of receiving a score greater than 750 on a GMAT test that has a mean of 476 and a standard deviation of 107? i.e., $P(X \geq 750) = ?$

This problem is asking for determining the area of the upper tail of the distribution. The Z score is: $Z = (750 - 476)/107 = 2.56$. From the table, the probability for this Z score is 0.4948. This is the probability of a GMAT with a score between 476 and 750. The rule is that when we want to find the probability in either tail, we must subtract the table value from 0.50. Thus, the answer to this problem is: $0.5 - 0.4948 = 0.0052$ or 0.52%. Note that $P(X \geq 750)$ is the same as $P(X > 750)$, because, in continuous distribution, the area under an exact number such as $X=750$ is zero. The following figure shows a graphic representation of this problem.

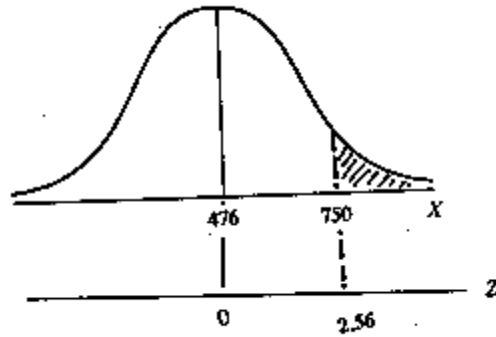


Figure 5

Question 3. What is the probability of receiving a score of 540 or less on a GMAT test that has a mean of 476 and a standard deviation of 107? i.e., $P(X \leq 540) = ?$. We are asked to determine the area under the curve for all values less than or equal to 540. The z score is: $z = (540 - 476) / 107 = 0.6$. From the table, the probability for this z score is 0.2257 which is the probability of getting a score between the mean (476) and 540. The rule is that when we want to find the probability between two values of x on either side of the mean, we just add the two areas together. Thus, the answer to this problem is: $0.5 + 0.2257 = 0.73$ or 73%. The following figure shows a graphic representation of this problem.

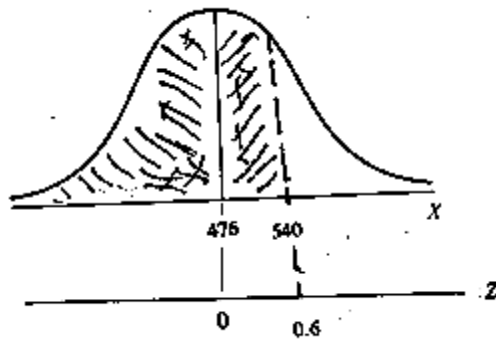


Figure 6

Question 4. What is the probability of receiving a score between 440 and 330 on a GMAT test that has a mean of 476 and a standard deviation of 107? i.e., $P(330 < 440) = ?$. The solution to this problem involves determining the area of the shaded slice in the lower half of the curve in

the following figure.

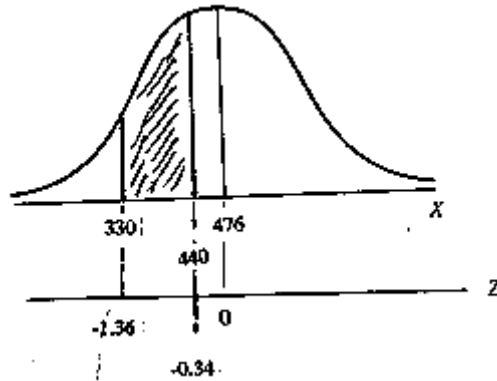
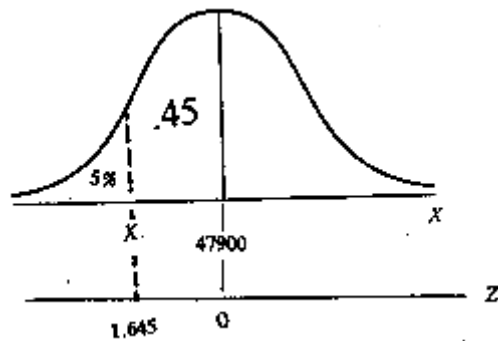


Figure 7

In this problem, the two values fall on the same side of the mean. The Z scores are: $Z_1 = (330 - 476)/107 = -1.36$, and $Z_2 = (440 - 476)/107 = -0.34$. The probability associated with $Z = -1.36$ is 0.4131, and the probability associated with $Z = -0.34$ is 0.1331. The rule is that when we want to find the probability between two values of X on one side of the mean, we just subtract the smaller area from the larger area to get the probability between the two values. Thus, the answer to this problem is: $0.4131 - 0.1331 = 0.28$ or 28%.

Example 2:

Suppose that a tire factory wants to set a mileage guarantee on its new model called LA 50 tire. Life tests indicated that the mean mileage is 47,900, and standard deviation of the normally distributed distribution of mileage is 2,050 miles. The factory wants to set the guaranteed mileage so that no more than 5% of the tires will have to be replaced. What guaranteed mileage should the factory announce? i.e., $P(X \leq ?) = 5\%$. In this problem, the mean and standard deviation are given, but X and Z are unknown. The problem is to solve for an X value that has 5% or 0.05 of the X values less than that value. If 0.05 of the values are less than X, then 0.45 lie between X and the mean ($0.5 - 0.05$), see the following graph.



Refer to the standard normal distribution table and search the body of the table for 0.45. Since the exact number is not found in the table, search for the closest number to 0.45. There are two values equidistant from 0.45-- 0.4505 and 0.4495. Move to the left from these values, and read the Z scores in the margin, which are: 1.65 and 1.64. Take the average of these two Z scores, i.e., $(1.65 + 1.64)/2 = 1.645$. Plug this number and the values of the mean and the standard deviation into the Z equation, you get:

$$Z = (X - \text{mean}) / \text{standard deviation} \text{ or } -1.645 = (X - 47,900) / 2,050 = 44,528 \text{ miles.}$$

Thus, the factory should set the guaranteed mileage at 44,528 miles if the objective is not to replace more than 5% of the tires.

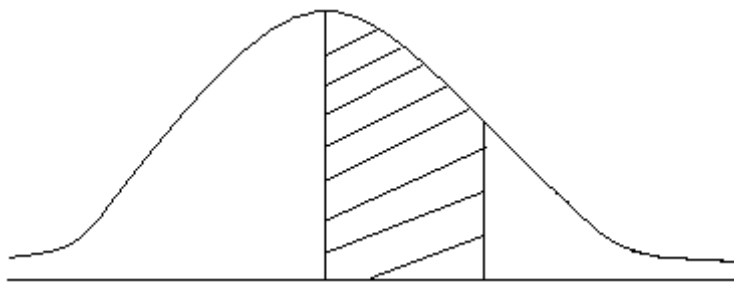
LECTURE 13

7.6 Finding z-scores from probabilities

This is more challenging, and requires you to use the table inversely. You must look up the area between zero and the value on the inside part of the table, and then read the z-score from the outside. Finally, decide if the z-score should be positive or negative, based on whether it was on the left side or the right side of the mean. Remember, z-scores can be negative, but areas or probabilities cannot be.

Situation	Instructions
Area between 0 and a value	Look up the area in the table Make negative if on the left side
Area in one tail	Subtract the area from 0.5000 Look up the difference in the table Make negative if in the left tail
Area including one complete half (Less than a positive or greater than a negative)	Subtract 0.5000 from the area Look up the difference in the table Make negative if on the left side
Within z units of the mean	Divide the area by 2 Look up the quotient in the table Use both the positive and negative z-scores
Two tails with equal area (More than z units from the mean)	Subtract the area from 1.000 Divide the area by 2 Look up the quotient in the table Use both the positive and negative z-scores

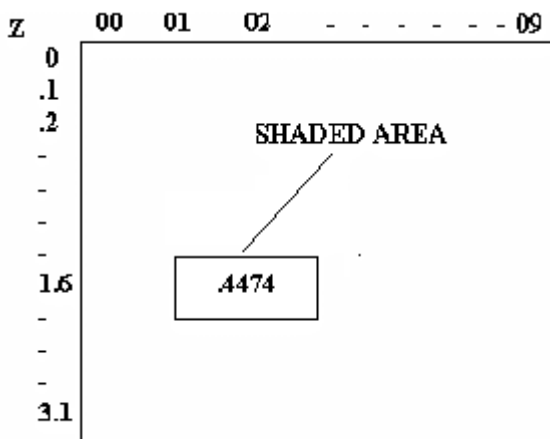
Using the table becomes proficient with practice, work lots of the normal probability problems!



-3 -2 -1 0 1 2 3 Z

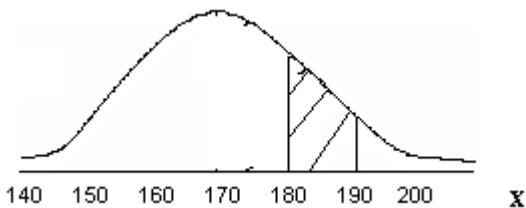
The shaded area between the middle ($Z=0$) and any value of Z is given in the tables. Assume $Z = 1.62$

The area from a Z of 0 to a Z of 1.62 is 0.4474. (Read from the tables)



The standard normal probability tables are at the back of this topic.

Example : The heights of adult males is normally distributed with mean 170cm and standard deviation 10cm.



Q1. Find the probability of a male between 180 and 190 cm.

Method:

a) Draw a diagram and shade the area required.

b) Use formula

$$Z_x = \frac{x - \text{mean}}{\text{std dev}}$$

Note: Strictly speaking a random variable is usually denoted by an uppercase **X** whereas a lower case **x** represents one of its values. You can show **x** instead of **X** if you wish.

To calculate **Z** values of the boundaries of shaded area

$$Z_x = \frac{x - \text{mean}}{\text{std dev}} = 1$$

$$Z_{190} = \frac{190 - 170}{10} = 2$$

Find areas in tables

Z	Area
1	.3413
2	.4772

d) Add or subtract table areas to get probability required.

$$P(180 \leq x \leq 190) = P(1 \leq Z \leq 2)$$

$$= 0.4772 - 0.3413$$

$$= 0.1359$$

Q2. Taller than 190cm

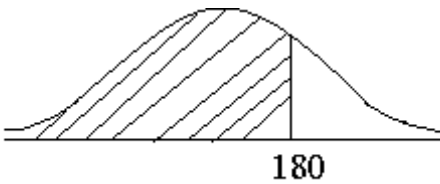
Tables : 0.4772



$$Z_{190} = \frac{190 - 170}{10} = 2$$

$$\begin{aligned} P(X > 190) &= 0.5 - .4772 \\ &= 0.0228 \end{aligned}$$

Q3. Shorter than 180cm



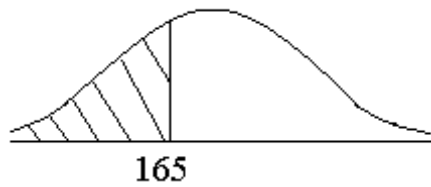
$$Z = \frac{180 - 170}{10} = 1$$

Tables: .3413

$$\begin{aligned} P(X < 180) &= 0.5 + 0.3413 \\ &= .8413 \end{aligned}$$

Q4. Shorter than 165cm.

Tables : 0 .1915



$$Z = \frac{165 - 170}{10} = -0.5$$

$$P(X < 165) = 0.5 - 0.1915 = 0.308$$

Chapter Review Questions

1. List 5 characteristics of the normal curve.

2. With the aid of clearly labeled normal curves, find the normal curve between

i) $Z = -2$ and $Z = 2.6$

ii) To the right of $Z = 2.4$

iii) To the left of $z = 2.0$

iv) Between $Z = 1.3$ and $Z = -1.2$

3. The marks obtained in a Business statistics CAT are normally distributed with mean 23 and standard deviation 4.2. Find the probability that a randomly selected student scores

i) More than 25 marks

ii) Between 20 and 25 marks

iii) Less than the mean mark?

4. The probability of any dry cell being defective is 15%. If 10 dry cells are selected at random, what is the probability that

- a) Fewer than 3 will be defective?*
- b) None will be defective?*
- c) Exactly 5 will be defective?*
- d) Between 6 and 8 will be defective?*
- e) At least 10 will not be defective?*

5. The average number of cars passing through a certain junction per minute is 10. What is the probability that in one minute,

- i) No fewer than 5 cars pass through the junction*
- ii) Exactly 12 cars pass through the junction*
- iii) Between 5 and 7 cars pass through the junction?*

Reference

- i. Essentials of statistics for Business and Economics by Anderson Sweetey Williams Pg 158-181, Pg191-212*
- ii. Quantitative Methods for Business by Donald Waters Pg382-393*

SAMPLE PAPERS

TEST PAPER 1

MT KENYA UNIVERSITY

UNIVERSITY EXAMINATIONS APRIL 2010

**FIRST YEAR SEMESTER I EXAMINATION FOR THE DEGREE OF BACHELOR OF
COMMERCE AND BACHELOR OF BUSINESS MANAGEMENT**

SBC223 / BBM223

INTRODUCTION TO BUSINESS STATISTICS

DATE: APRIL 2010

TIME: 3 HOURS

Instructions: Answer Question One and Any Other Two Questions

QUESTION ONE (Compulsory) (30 marks)

- (a) Using a well labeled diagram, show the position of the mean, mode and the median in a negatively skewed distribution (3marks)
- (b) (b) The weighted mean of the two numbers 30 and 15 is 20. If the weightings are 2 and x , find x . (3marks)
- (c) For a skewed distribution, the mean is 86, the median is 20 and the standard deviation is 5. Calculate the Pearson's coefficient of skewness and sketch the curve (3marks)

- (d) State three properties of the normal distribution. (3marks)
- (e) Given a set of data; 2,9,8,4,7,6
- i) Calculate the arithmetic mean (2 marks)
 - ii) Calculate the geometric mean (2 marks)
 - iii) Calculate the harmonic mean (2 marks)
 - iv) State the median (1 mark)
 - v) Calculate the standard deviation. (3 marks)
- (f) Cartons of orange juice are advertised as containing 1litre. A random sample of 100 cartons gave the following results for the volume, x.
- $$\sum x = 101.4, \quad \sum x^2 = 102.83 .$$
- Calculate the mean and standard deviation of the volume of juice in those 100 cartons. (3marks)
- ((f) Indicate the level of measurement in each of the following:
- i) Marital status of a group of senior citizens. (1mark)
 - ii) Level of satisfaction with a meal at a family restaurant. (1mark)
 - iii) Rank of officers in an army regiment. (1mark)
 - iv) The weight of dog food. (1mark)
 - v) Time spent in a park. (1mark)

QUESTION TWO **(20 marks)**

The manager of a fast food restaurant is concerned that the customers are waiting for too long for their food. She decides to gather some statistics on customer waiting times and the following times (in minutes) are recorded).

1.25 2.5 8.5 4.6 10.5 3.4 3.7 6.25 7.7 4.1 5.15 5.95 7.35 5.8
2.9 3.4 6.6 8.8 2.7 10.2 4.5 5.2 4.1 2.5 3.8 2.1 5.5 6.25
4.3 1.8 3.7 4.4 6.2 3.3 7.2 8.6 3.45 6.55 2.85 9.4 4.25 5.6
11.9 6.4 4.8 5.8 2.5 4.1 8.1 6.1

- (a) Group this
data into classes of 0 - 1.9 , 2 - 3.9 etc and construct a frequency table which also shows cumulative frequencies. (4marks)
- (b) Construct a histogram for the data. (4marks)
- (c) Construct a less than ogive for the data and answer the following questions (4marks)
- i) How many customers have to wait for less than 4 minutes to be served? (2marks)
- ii) What percentage of customers has to wait for less than 5 minutes for their food? (2marks)
- iii) If the restaurant's goal is for 90% of the customers to be given their food within 8 minutes, are they achieving this goal? (2marks)
- iv) What is the mean waiting time? 2marks)

QUESTION THREE

(20 marks)

- a) A random sample of 51 people was asked to record the number of miles they travelled by car in a given week. The distances to the nearest mile, are shown below.

42	93	46	52	72	77	53	41	48	86
62	54	85	60	58	43	58	43	58	74
52	82	78	86	94	63	72	63	72	44
78	56	80	44	52	74	68	82	57	47

- a) Construct a stem and leaf diagram to represent these data. (4marks)
- b) Find the median and the quartiles of this distribution. (4marks)
- c) Draw a box plot to represent these data. (4marks)
- d) Give one advantage of using
- i) a stem and leaf diagram (2marks)
 - ii) a box plot (2marks)
- to illustrate data such as the one given above.
- b) At the end of a statistics course, Diana sits for two written papers, P1 and P2 and hands in a piece of course work. Her marks out 100 were 76 for P1 and 67 for P2 and she gained 81 marks for her course work. Her overall percentage is to be weighted so that the two written papers account for 40% while the course work accounts for 20%. Calculate Diana's overall percentage mark. (4 marks)

QUESTION FOUR

(20 marks)

- a) A reaction time experiment was performed first with 21 girls, and then with 24 boys. The results are shown on the stem and leaf diagram below

Key(Girls)

Key(Boys)

6|1 means 0.16 sec

1|8 means 0.18 sec

Girls						Boys				
				4	2	4				
		3	3	2	2	2	2			
			1	0	0	2	0	0	0	1 1
9	9	9	8	8	1	8	8	8		
			7	7	6	1	6	6	7	7
5	5	4	4	4	1	4	5	5		
					1	2	3			
					1	0	1	1		
				0	0	9				

- i) Find the median and inter quartile range for both sets of reaction times (6marks)
 ii) Comment on your answers. (2marks)

- b) In a practical class, students timed how long it took for a sample of their saliva to break down a 2% starch solution. The times to the nearest second are shown below:

<u>Time (seconds)</u>	<u>frequency</u>
11 - 20	1
21 - 30	2
31 - 40	5
41 - 50	11
51 - 60	8
61 - 70	2
71 - 80	1

- i) Calculate the mean and standard deviation for this data. (4marks)
- ii) Calculate the inter quartile range (3marks)
- iii) Calculate the coefficient of skewness (3marks)
- iv) Comment on the skewness of this distribution (2marks)

QUESTION FOUR (20 marks)

- c) A random sample of 51 people was asked to record the number of miles they travelled by car in a given week. The distances to the nearest mile, are shown below.

42	93	46	52	72	77	53	41	48	86
62	54	85	60	58	43	58	43	58	74
52	82	78	86	94	63	72	63	72	44
78	56	80	44	52	74	68	82	57	47

- e) Construct a stem and leaf diagram to represent these data. (4marks)
- f) Find the median and the quartiles of this distribution. (4marks)
- g) Draw a box plot to represent these data. (4marks)
- h) Give one advantage of using
- iii) a stem and leaf diagram (2marks)
- iv) a box plot (2marks)
- to illustrate data such as the one given above.
- d) At the end of a statistics course, Diana sits for two written papers, P1 and P2 and hands in a piece of course work. Her marks out 100 were 76 for P1 and 67 for P2 and she gained 81 marks for her course work. Her overall percentage is to be weighted so that the two written papers account for 40% while the course work accounts for 20%. Calculate Diana's overall percentage mark. (4 marks)

QUESTION FIVE (20 Marks)

- a) A shopkeeper wanted to investigate whether or not there was a correlation between the prices of food 10 years ago in 1992, with their prices today. He chose 8 everyday items and the prices are given in the table below.

	sugar	milk	eggs	rolls	tea bags	coffee	potatoes	flour
1992 price	\$ 1.44	\$ 0.80	\$ 2.16	\$ 1.80	\$ 0.92	\$ 3.16	\$ 1.32	\$ 1.12
2002 price	\$ 2.20	\$ 1.04	\$ 2.64	\$ 3.00	\$ 1.32	\$ 2.28	\$ 1.92	\$ 1.44

- (a) Calculate the mean and the standard deviation of the prices
- (i) in 1992;
- (ii) in 2002. **(4)**
- (b) (i) Given that $s_{xy} = 0.3104$, calculate the correlation coefficient.
- (ii) Comment on the relationship between the prices. **(4)**
- (c) Find the equation of the line of the best fit in the form $y = mx + c$. **(3)**
- (d) What would you expect to pay now for an item costing \$2.60 in 1992? **(1)**
- (e) Which item would you omit to increase the correlation coefficient? **(2)**
- b) For a skewed distribution, the mean is 86, the median is 20 and the standard deviation is 5. Calculate the Pearson's coefficient of skewness and sketch the curve. **(2)**
- c) Briefly explain any two instances in which the knowledge of Business Statistics may be applied in making a managerial decision. **(4)**

TEST PAPER 2

MT KENYA UNIVERSITY

UNIVERSITY EXAMINATIONS 2009/2010

**FIRST YEAR SEMESTER I EXAMINATION FOR THE DEGREE OF BACHELOR OF
COMMERCE AND BACHELOR OF BUSINESS MANAGEMENT**

SBC223 / BBM22

: INTRODUCTION TO BUSINESS STATISTICS

DATE: JANUARY 2010

TIME: 2 HOURS

Instructions: Answer Question One and Any Other Two Questions

QUESTION ONE (Compulsory) (30 marks)

- a) What is Business statistics? (2 marks)
- b) Differentiate between descriptive and inferential statistics. (2 marks)
- c) State Three reasons why a researcher may prefer to study a sample instead of the whole population. (3 marks)
- d) Given a set of data; 2,9,8,3,5
 - vi) Calculate the arithmetic mean (2 marks)
 - vii) Calculate the geometric mean (2 marks)
 - viii) Calculate the harmonic mean (2 marks)
 - ix) State the median (1 mark)
 - x) Calculate the standard deviation. (3 marks)
- e) Using a clearly labeled diagram, show the position of the mode, mean and median of a positively skewed distribution. (2 marks)

f) In a certain set of data, the following measures were found:

Minimum value = 13

$Q_1 = 15$

Median = 18

$Q_3 = 22$

Maximum value = 30

- i) Construct a box plot to represent the above information. (3marks)
- ii) Comment on the skewness of the data involved. (1 mark)
- g) State the Chebyshev's theorem. (3marks)

h) Equity Bank Ltd is studying the number of times their automatic teller machine located at Tom Mboya Street is used daily. The following is a list of the number of times the teller was used during the last 30 days.

83 64 84 76 54 84 75 59 70 61

63 83 84 70 68 52 65 90 52 77

95 36 78 61 59 84 95 87 47 60

Construct a stem and leaf chart to represent the information above. (4 marks)

QUESTION TWO (20 mark

- a) The following frequency distribution shows the daily milk deliveries for Maziwa Farmers' Union.

Amt of milk(ltrs)	6–10	10 –14	14 –18	18–22	22–26	26–30	30 –34
No. of farmers	5	8	12	10	8	2	4

34–38	38-42
0	1

- i) Construct a **relative frequency** histogram for this information (4 marks)
- ii) How many farmers delivered at least 22 litres of milk daily? (2 marks)
- iii) What percent of farmers delivered less than 10 litres of milk daily? (2 marks)
- iv) What proportion of farmers delivered less than 10 litres of milk daily? (2 marks)

b) A research was carried out to determine whether there is a relation between absenteeism and passing the exam. The table below shows some results.

Number of days absent	Pass	Fail	TOTAL
Less than 4 days	135	110	
Less than 7 days	36	4	
Less than 45 days	9	6	

- c) Calculate the probability that a student selected at random:
- i) Was absent for less than four days (2 marks)
- ii) Passed the exam (2marks)
- ii) Was absent for less than seven days. (2 marks)
- iii) Passed given that he/she was absent for less than four days. (2 marks)

iv) Failed given that he/she was absent for less than seven days. (2 marks)

d) Differentiate between nominal and ordinal scales of measurement. (2 marks)

QUESTION THREE (20 marks)

a) Given the set of data

2, 3, 7, 4, 9

Calculate the **sample** standard deviation. (4 marks)

b) The variation of incomes of executives is to be compared with the variation of incomes of unskilled employees. For a sample of executives, mean = Ksh 500,000 and Standard deviation = Ksh 50,000. For a sample of unskilled employees, the mean = Ksh 12,000 and standard deviation = Ksh 2000.

Is the variation among the executives greater than the variation among the unskilled employees? Justify your answer. (5 marks)

c) At the end of a statistics course, Diana sits for two written papers, P1 and P2 and hands in a piece of course work. Her marks out 100 were 76 for P1 and 67 for P2 and she gained 81 marks for her course work. Her overall percentage is to be weighted so that the two written papers account for 40% while the course work accounts for 20%. Calculate Diana's overall percentage mark. (4 marks)

d) What is a discrete variable? (2 marks)

e) Dispersion is a statistical name for spread or variability while skewness describes how non-symmetric the data is. True or False? (2 marks)

f) State three characteristics of the median as a measure of central tendency. (3 marks)

QUESTION FOUR**(20 marks)**

a) Marks out of 100 for 100 students were tabulated as shown below:

Marks	f
11 – 20	4
21 – 30	16
31 – 40	27
41 – 50	32
51 – 60	15
61 – 70	4
71 – 80	2

- i) Calculate the mean mark. (4 marks)
- ii) Draw an ogive to represent the above data. (4 marks)
- iii) Use the ogive to estimate the upper quartile and the lower quartile (2 marks)
- iii) Estimate from your ogive the minimum marks required to get grade A if
- iv) only 5 students are to get grade A. (2 marks)
- v) How many students were to pass if the pass mark was set at 25 marks? (2 marks)

b) A group of students has measured the heights of 90 trees. The class calculate the mean height to be $\bar{x} = 12.4$ m with standard deviation $s = 5.35$ m. One student notices that two of the measurements, 44.5 m and 43.2 m, are much too big and must be wrong.

- (a) How many standard deviations away from the mean of 12.4 is the value 44.5? (3marks)

The incorrect measurements of 44.5 m and 43.2 m must be removed from the data.

(b) Calculate the new value of x after removing the two unwanted values.

(3 marks)

QUESTION FIVE **(20marks)**

The following data relates to daily bill on consumption of a certain commodity for 60 households

Daily bills(KSh)	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of households	6	7	11	10	6	5	9	3	3

i) Calculate the mean (4 marks)

ii) Calculate the median (3marks)

iii) Calculate the standard deviation (3marks)

iv) Calculate the coefficient of skewness (4 marks)

v) Comment on the skewness of this distribution (2 marks)

vi) Calculate the coefficient of variation (4 marks)

APPENDIX

I.NORMAL DISTRIBUTION TABLES

2. BINOMIAL TABLES

3.POISSON TABLES

Standard Normal Probabilities

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633

1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

The values in the table are the areas between zero and the z-score. That is, $P(0 < Z < z\text{-score})$

Binomial Distribution Tables

N = 2

K \ P=.1 .2 .3 .4 .5 .6 .7 .8 .9

0	0.81	0.64	0.49	0.36	0.25	0.16	0.09	0.04	0.01
1	0.99	0.96	0.91	0.84	0.75	0.64	0.51	0.36	0.19
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

N = 3

K \ P=.1 .2 .3 .4 .5 .6 .7 .8 .9

0	0.729	0.512	0.343	0.216	0.125	0.064	0.027	0.008	0.001
1	0.972	0.896	0.784	0.648	0.500	0.352	0.216	0.104	0.028
2	0.999	0.992	0.973	0.936	0.875	0.784	0.657	0.488	0.271
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

N = 4

K \ P=.1 .2 .3 .4 .5 .6 .7 .8 .9

0	0.6561	0.4096	0.2401	0.1296	0.0625	0.0256	0.0081	0.0016	0.0001
1	0.9477	0.8192	0.6517	0.4752	0.3125	0.1792	0.0837	0.0272	0.0037
2	0.9963	0.9728	0.9163	0.8208	0.6875	0.5248	0.3483	0.1808	0.0523
3	0.9999	0.9984	0.9919	0.9744	0.9375	0.8704	0.7599	0.5904	0.3439
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

N = 5

K \ P=.1 .2 .3 .4 .5 .6 .7 .8 .9

0	0.59049	0.32768	0.16807	0.07776	0.03125	0.01024	0.00243	0.00032	0.00001
1	0.91854	0.73728	0.52822	0.33696	0.18750	0.08704	0.03078	0.00672	0.00046
2	0.99144	0.94208	0.83692	0.68256	0.50000	0.31744	0.16308	0.05792	0.00856
3	0.99954	0.99328	0.96922	0.91296	0.81250	0.66304	0.47178	0.26272	0.08146
4	0.99999	0.99968	0.99757	0.98976	0.96875	0.92224	0.83193	0.67232	0.40951
5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

N = 6

K \ P=.1 .2 .3 .4 .5 .6 .7 .8 .9

0	0.53144	0.26214	0.11765	0.04666	0.01562	0.00410	0.00073	0.00006	0.00000
1	0.88574	0.65536	0.42018	0.23328	0.10938	0.04096	0.01094	0.00160	0.00006
2	0.98415	0.90112	0.74431	0.54432	0.34375	0.17920	0.07047	0.01696	0.00127

3		0.99873	0.98304	0.92953	0.82080	0.65625	0.45568	0.25569	0.09888	0.01585
4		0.99994	0.99840	0.98906	0.95904	0.89062	0.76672	0.57982	0.34464	0.11426
5		1.00000	0.99994	0.99927	0.99590	0.98438	0.95334	0.88235	0.73786	0.46856
6		1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

N = 7

K \ P=.	1	.2	.3	.4	.5	.6	.7	.8	.9
---------	---	----	----	----	----	----	----	----	----

0		0.47830	0.20972	0.08235	0.02799	0.00781	0.00164	0.00022	0.00001	0.00000
1		0.85031	0.57672	0.32942	0.15863	0.06250	0.01884	0.00379	0.00037	0.00001
2		0.97431	0.85197	0.64707	0.41990	0.22656	0.09626	0.02880	0.00467	0.00018
3		0.99727	0.96666	0.87396	0.71021	0.50000	0.28979	0.12604	0.03334	0.00273
4		0.99982	0.99533	0.97120	0.90374	0.77344	0.58010	0.35293	0.14803	0.02569
5		0.99999	0.99963	0.99621	0.98116	0.93750	0.84137	0.67058	0.42328	0.14969
6		1.00000	0.99999	0.99978	0.99836	0.99219	0.97201	0.91765	0.79028	0.52170
7		1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

N = 8

K \ P=.	1	.2	.3	.4	.5	.6	.7	.8	.9
---------	---	----	----	----	----	----	----	----	----

0		0.43047	0.16777	0.05765	0.01680	0.00391	0.00066	0.00007	0.00000	0.00000
1		0.81310	0.50332	0.25530	0.10638	0.03516	0.00852	0.00129	0.00008	0.00000
2		0.96191	0.79692	0.55177	0.31539	0.14453	0.04981	0.01129	0.00123	0.00002
3		0.99498	0.94372	0.80590	0.59409	0.36328	0.17367	0.05797	0.01041	0.00043
4		0.99957	0.98959	0.94203	0.82633	0.63672	0.40591	0.19410	0.05628	0.00502
5		0.99998	0.99877	0.98871	0.95019	0.85547	0.68461	0.44823	0.20308	0.03809
6		1.00000	0.99992	0.99871	0.99148	0.96484	0.89362	0.74470	0.49668	0.18690
7		1.00000	1.00000	0.99993	0.99934	0.99609	0.98320	0.94235	0.83223	0.56953
8		1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

N = 9

K \ P=.	1	.2	.3	.4	.5	.6	.7	.8	.9
---------	---	----	----	----	----	----	----	----	----

0		0.38742	0.13422	0.04035	0.01008	0.00195	0.00026	0.00002	0.00000	0.00000
1		0.77484	0.43621	0.19600	0.07054	0.01953	0.00380	0.00043	0.00002	0.00000
2		0.94703	0.73820	0.46283	0.23179	0.08984	0.02503	0.00429	0.00031	0.00000
3		0.99167	0.91436	0.72966	0.48261	0.25391	0.09935	0.02529	0.00307	0.00006
4		0.99911	0.98042	0.90119	0.73343	0.50000	0.26657	0.09881	0.01958	0.00089
5		0.99994	0.99693	0.97471	0.90065	0.74609	0.51739	0.27034	0.08564	0.00833
6		1.00000	0.99969	0.99571	0.97497	0.91016	0.76821	0.53717	0.26180	0.05297
7		1.00000	0.99998	0.99957	0.99620	0.98047	0.92946	0.80400	0.56379	0.22516
8		1.00000	1.00000	0.99998	0.99974	0.99805	0.98992	0.95965	0.86578	0.61258
9		1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

N=10

K \ P= .1 .2 .3 .4 .5 .6 .7 .8 .9

0	0.34868	0.10737	0.02825	0.00605	0.00098	0.00010	0.00001	0.00000	0.00000
1	0.73610	0.37581	0.14931	0.04636	0.01074	0.00168	0.00014	0.00000	0.00000
2	0.92981	0.67780	0.38278	0.16729	0.05469	0.01229	0.00159	0.00008	0.00000
3	0.98720	0.87913	0.64961	0.38228	0.17188	0.05476	0.01059	0.00086	0.00001
4	0.99837	0.96721	0.84973	0.63310	0.37695	0.16624	0.04735	0.00637	0.00015
5	0.99985	0.99363	0.95265	0.83376	0.62305	0.36690	0.15027	0.03279	0.00163
6	0.99999	0.99914	0.98941	0.94524	0.82812	0.61772	0.35039	0.12087	0.01280
7	1.00000	0.99992	0.99841	0.98771	0.94531	0.83271	0.61722	0.32220	0.07019
8	1.00000	1.00000	0.99986	0.99832	0.98926	0.95364	0.85069	0.62419	0.26390
9	1.00000	1.00000	0.99999	0.99990	0.99902	0.99395	0.97175	0.89263	0.65132
10	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Cumulative Binomial Distribution Table

n=1

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.9900	0.9800	0.9700	0.9600	0.9500	0.9400	0.9300	0.9200	0.9100
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.4500	0.4000	0.3500	0.3000	0.2500	0.2000	0.1500	0.1000	0.0900
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0800	0.0700	0.0600	0.0500	0.0400	0.0300	0.0200	0.0100	0.0000
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=2

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.9801	0.9604	0.9409	0.9216	0.9025	0.8836	0.8649	0.8464	0.8281
1	0.9999	0.9996	0.9991	0.9984	0.9975	0.9964	0.9951	0.9936	0.9919
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
1	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.2025	0.1600	0.1225	0.0900	0.0625	0.0400	0.0225	0.0100	0.0081
1	0.6975	0.6400	0.5775	0.5100	0.4375	0.3600	0.2775	0.1900	0.1719
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0064	0.0049	0.0036	0.0025	0.0016	0.0009	0.0004	0.0001	0.0000
1	0.1536	0.1351	0.1164	0.0975	0.0784	0.0591	0.0396	0.0199	0.0000

2 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000

n=3

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.9703	0.9412	0.9127	0.8847	0.8574	0.8306	0.8044	0.7787	0.7536
1	0.9997	0.9988	0.9974	0.9953	0.9928	0.9896	0.9860	0.9818	0.9772
2	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995	0.9993
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
1	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000
2	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0911	0.0640	0.0429	0.0270	0.0156	0.0080	0.0034	0.0010	0.0007
1	0.4253	0.3520	0.2818	0.2160	0.1563	0.1040	0.0608	0.0280	0.0228
2	0.8336	0.7840	0.7254	0.6570	0.5781	0.4880	0.3859	0.2710	0.2464
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0005	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
1	0.0182	0.0140	0.0104	0.0073	0.0047	0.0026	0.0012	0.0003	0.0000

2	0.2213	0.1956	0.1694	0.1426	0.1153	0.0873	0.0588	0.0297	0.0000
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=4

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.9606	0.9224	0.8853	0.8493	0.8145	0.7807	0.7481	0.7164	0.6857
1	0.9994	0.9977	0.9948	0.9909	0.9860	0.9801	0.9733	0.9656	0.9570
2	1.0000	1.0000	0.9999	0.9998	0.9995	0.9992	0.9987	0.9981	0.9973
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
1	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
2	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
3	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0410	0.0256	0.0150	0.0081	0.0039	0.0016	0.0005	0.0001	0.0001
1	0.2415	0.1792	0.1265	0.0837	0.0508	0.0272	0.0120	0.0037	0.0027
2	0.6090	0.5248	0.4370	0.3483	0.2617	0.1808	0.1095	0.0523	0.0430
3	0.9085	0.8704	0.8215	0.7599	0.6836	0.5904	0.4780	0.3439	0.3143
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0019	0.0013	0.0008	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000
2	0.0344	0.0267	0.0199	0.0140	0.0091	0.0052	0.0023	0.0006	0.0000
3	0.2836	0.2519	0.2193	0.1855	0.1507	0.1147	0.0776	0.0394	0.0000
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=5

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.9510	0.9039	0.8587	0.8154	0.7738	0.7339	0.6957	0.6591	0.6240
1	0.9990	0.9962	0.9915	0.9852	0.9774	0.9681	0.9575	0.9456	0.9326
2	1.0000	0.9999	0.9997	0.9994	0.9988	0.9980	0.9969	0.9955	0.9937
3	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
1	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
2	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
3	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
4	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0185	0.0102	0.0053	0.0024	0.0010	0.0003	0.0001	0.0000	0.0000
1	0.1312	0.0870	0.0540	0.0308	0.0156	0.0067	0.0022	0.0005	0.0003
2	0.4069	0.3174	0.2352	0.1631	0.1035	0.0579	0.0266	0.0086	0.0063
3	0.7438	0.6630	0.5716	0.4718	0.3672	0.2627	0.1648	0.0815	0.0674
4	0.9497	0.9222	0.8840	0.8319	0.7627	0.6723	0.5563	0.4095	0.3760
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0045	0.0031	0.0020	0.0012	0.0006	0.0003	0.0001	0.0000	0.0000
3	0.0544	0.0425	0.0319	0.0226	0.0148	0.0085	0.0038	0.0010	0.0000
4	0.3409	0.3043	0.2661	0.2262	0.1846	0.1413	0.0961	0.0490	0.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=6

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.9415	0.8858	0.8330	0.7828	0.7351	0.6899	0.6470	0.6064	0.5679
1	0.9985	0.9943	0.9875	0.9784	0.9672	0.9541	0.9392	0.9227	0.9048
2	1.0000	0.9998	0.9995	0.9988	0.9978	0.9962	0.9942	0.9915	0.9882
3	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9992
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
1	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
2	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
3	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
4	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
5	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0083	0.0041	0.0018	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000
1	0.0692	0.0410	0.0223	0.0109	0.0046	0.0016	0.0004	0.0001	0.0000
2	0.2553	0.1792	0.1174	0.0705	0.0376	0.0170	0.0059	0.0013	0.0008
3	0.5585	0.4557	0.3529	0.2557	0.1694	0.0989	0.0473	0.0159	0.0118
4	0.8364	0.7667	0.6809	0.5798	0.4661	0.3446	0.2235	0.1143	0.0952
5	0.9723	0.9533	0.9246	0.8824	0.8220	0.7379	0.6229	0.4686	0.4321
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

2	0.0005	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0085	0.0058	0.0038	0.0022	0.0012	0.0005	0.0002	0.0000	0.0000
4	0.0773	0.0608	0.0459	0.0328	0.0216	0.0125	0.0057	0.0015	0.0000
5	0.3936	0.3530	0.3101	0.2649	0.2172	0.1670	0.1142	0.0585	0.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=7

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.9321	0.8681	0.8080	0.7514	0.6983	0.6485	0.6017	0.5578	0.5168
1	0.9980	0.9921	0.9829	0.9706	0.9556	0.9382	0.9187	0.8974	0.8745
2	1.0000	0.9997	0.9991	0.9980	0.9962	0.9937	0.9903	0.9860	0.9807
3	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9988	0.9982
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
1	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
2	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
3	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
4	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
5	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
6	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922

7 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0037	0.0016	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
1	0.0357	0.0188	0.0090	0.0038	0.0013	0.0004	0.0001	0.0000	0.0000
2	0.1529	0.0963	0.0556	0.0288	0.0129	0.0047	0.0012	0.0002	0.0001
3	0.3917	0.2898	0.1998	0.1260	0.0706	0.0333	0.0121	0.0027	0.0018
4	0.6836	0.5801	0.4677	0.3529	0.2436	0.1480	0.0738	0.0257	0.0193
5	0.8976	0.8414	0.7662	0.6706	0.5551	0.4233	0.2834	0.1497	0.1255
6	0.9848	0.9720	0.9510	0.9176	0.8665	0.7903	0.6794	0.5217	0.4832
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0012	0.0007	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0140	0.0097	0.0063	0.0038	0.0020	0.0009	0.0003	0.0000	0.0000
5	0.1026	0.0813	0.0618	0.0444	0.0294	0.0171	0.0079	0.0020	0.0000
6	0.4422	0.3983	0.3515	0.3017	0.2486	0.1920	0.1319	0.0679	0.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=8

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
---	---------------------	---------------------	---------------------	---------------------	---------------------	---------------------	---------------------	---------------------	---------------------

0	0.9227	0.8508	0.7837	0.7214	0.6634	0.6096	0.5596	0.5132	0.4703
1	0.9973	0.9897	0.9777	0.9619	0.9428	0.9208	0.8965	0.8702	0.8423
2	0.9999	0.9996	0.9987	0.9969	0.9942	0.9904	0.9853	0.9789	0.9711
3	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9987	0.9978	0.9966
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
1	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
2	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
3	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
4	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
5	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
6	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0017	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0181	0.0085	0.0036	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000
2	0.0885	0.0498	0.0253	0.0113	0.0042	0.0012	0.0002	0.0000	0.0000

3	0.2604	0.1737	0.1061	0.0580	0.0273	0.0104	0.0029	0.0004	0.0003
4	0.5230	0.4059	0.2936	0.1941	0.1138	0.0563	0.0214	0.0050	0.0034
5	0.7799	0.6846	0.5722	0.4482	0.3215	0.2031	0.1052	0.0381	0.0289
6	0.9368	0.8936	0.8309	0.7447	0.6329	0.4967	0.3428	0.1869	0.1577
7	0.9916	0.9832	0.9681	0.9424	0.8999	0.8322	0.7275	0.5695	0.5297
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0022	0.0013	0.0007	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000
5	0.0211	0.0147	0.0096	0.0058	0.0031	0.0013	0.0004	0.0001	0.0000
6	0.1298	0.1035	0.0792	0.0572	0.0381	0.0223	0.0103	0.0027	0.0000
7	0.4868	0.4404	0.3904	0.3366	0.2786	0.2163	0.1492	0.0773	0.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=9

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.9135	0.8337	0.7602	0.6925	0.6302	0.5730	0.5204	0.4722	0.4279
1	0.9966	0.9869	0.9718	0.9522	0.9288	0.9022	0.8729	0.8417	0.8088
2	0.9999	0.9994	0.9980	0.9955	0.9916	0.9862	0.9791	0.9702	0.9595
3	1.0000	1.0000	0.9999	0.9997	0.9994	0.9987	0.9977	0.9963	0.9943
4	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995

5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
1	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
2	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
3	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
4	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
5	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
6	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0008	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0091	0.0038	0.0014	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000
2	0.0498	0.0250	0.0112	0.0043	0.0013	0.0003	0.0000	0.0000	0.0000
3	0.1658	0.0994	0.0536	0.0253	0.0100	0.0031	0.0006	0.0001	0.0000
4	0.3786	0.2666	0.1717	0.0988	0.0489	0.0196	0.0056	0.0009	0.0005
5	0.6386	0.5174	0.3911	0.2703	0.1657	0.0856	0.0339	0.0083	0.0057

6	0.8505	0.7682	0.6627	0.5372	0.3993	0.2618	0.1409	0.0530	0.0405
7	0.9615	0.9295	0.8789	0.8040	0.6997	0.5638	0.4005	0.2252	0.1912
8	0.9954	0.9899	0.9793	0.9596	0.9249	0.8658	0.7684	0.6126	0.5721
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0037	0.0023	0.0013	0.0006	0.0003	0.0001	0.0000	0.0000	0.0000
6	0.0298	0.0209	0.0138	0.0084	0.0045	0.0020	0.0006	0.0001	0.0000
7	0.1583	0.1271	0.0978	0.0712	0.0478	0.0282	0.0131	0.0034	0.0000
8	0.5278	0.4796	0.4270	0.3698	0.3075	0.2398	0.1663	0.0865	0.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=10

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.9044	0.8171	0.7374	0.6648	0.5987	0.5386	0.4840	0.4344	0.3894
1	0.9957	0.9838	0.9655	0.9418	0.9139	0.8824	0.8483	0.8121	0.7746
2	0.9999	0.9991	0.9972	0.9938	0.9885	0.9812	0.9717	0.9599	0.9460
3	1.0000	1.0000	0.9999	0.9996	0.9990	0.9980	0.9964	0.9942	0.9912
4	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994	0.9990
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
1	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
2	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
3	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
4	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
5	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
6	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
7	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0045	0.0017	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0274	0.0123	0.0048	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3	0.1020	0.0548	0.0260	0.0106	0.0035	0.0009	0.0001	0.0000	0.0000
4	0.2616	0.1662	0.0949	0.0473	0.0197	0.0064	0.0014	0.0001	0.0001

5	0.4956	0.3669	0.2485	0.1503	0.0781	0.0328	0.0099	0.0016	0.0010
6	0.7340	0.6177	0.4862	0.3504	0.2241	0.1209	0.0500	0.0128	0.0088
7	0.9004	0.8327	0.7384	0.6172	0.4744	0.3222	0.1798	0.0702	0.0540
8	0.9767	0.9536	0.9140	0.8507	0.7560	0.6242	0.4557	0.2639	0.2254
9	0.9975	0.9940	0.9865	0.9718	0.9437	0.8926	0.8031	0.6513	0.6106
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0006	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0058	0.0036	0.0020	0.0010	0.0004	0.0001	0.0000	0.0000	0.0000
7	0.0401	0.0283	0.0188	0.0115	0.0062	0.0028	0.0009	0.0001	0.0000
8	0.1879	0.1517	0.1176	0.0861	0.0582	0.0345	0.0162	0.0043	0.0000
9	0.5656	0.5160	0.4614	0.4013	0.3352	0.2626	0.1829	0.0956	0.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=11

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.8953	0.8007	0.7153	0.6382	0.5688	0.5063	0.4501	0.3996	0.3544
1	0.9948	0.9805	0.9587	0.9308	0.8981	0.8618	0.8228	0.7819	0.7399
2	0.9998	0.9988	0.9963	0.9917	0.9848	0.9752	0.9630	0.9481	0.9305

3	1.0000	1.0000	0.9998	0.9993	0.9984	0.9970	0.9947	0.9915	0.9871
4	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9995	0.9990	0.9983
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
1	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0139	0.0059
2	0.9104	0.7788	0.6174	0.4552	0.3127	0.2001	0.1189	0.0652	0.0327
3	0.9815	0.9306	0.8389	0.7133	0.5696	0.4256	0.2963	0.1911	0.1133
4	0.9972	0.9841	0.9496	0.8854	0.7897	0.6683	0.5328	0.3971	0.2744
5	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7535	0.6331	0.5000
6	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8262	0.7256
7	1.0000	1.0000	0.9998	0.9988	0.9957	0.9878	0.9707	0.9390	0.8867
8	1.0000	1.0000	1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673
9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9941
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
----------	---------------------	---------------------	---------------------	---------------------	---------------------	---------------------	---------------------	---------------------	---------------------

0	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0022	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0148	0.0059	0.0020	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000
3	0.0610	0.0293	0.0122	0.0043	0.0012	0.0002	0.0000	0.0000	0.0000
4	0.1738	0.0994	0.0501	0.0216	0.0076	0.0020	0.0003	0.0000	0.0000
5	0.3669	0.2465	0.1487	0.0782	0.0343	0.0117	0.0027	0.0003	0.0002
6	0.6029	0.4672	0.3317	0.2103	0.1146	0.0504	0.0159	0.0028	0.0017
7	0.8089	0.7037	0.5744	0.4304	0.2867	0.1611	0.0694	0.0185	0.0129
8	0.9348	0.8811	0.7999	0.6873	0.5448	0.3826	0.2212	0.0896	0.0695
9	0.9861	0.9698	0.9394	0.8870	0.8029	0.6779	0.5078	0.3026	0.2601
10	0.9986	0.9964	0.9912	0.9802	0.9578	0.9141	0.8327	0.6862	0.6456
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0010	0.0005	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0085	0.0053	0.0030	0.0016	0.0007	0.0002	0.0000	0.0000	0.0000
8	0.0519	0.0370	0.0248	0.0152	0.0083	0.0037	0.0012	0.0002	0.0000
9	0.2181	0.1772	0.1382	0.1019	0.0692	0.0413	0.0195	0.0052	0.0000
10	0.6004	0.5499	0.4937	0.4312	0.3618	0.2847	0.1993	0.1047	0.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=12

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.8864	0.7847	0.6938	0.6127	0.5404	0.4759	0.4186	0.3677	0.3225
1	0.9938	0.9769	0.9514	0.9191	0.8816	0.8405	0.7967	0.7513	0.7052
2	0.9998	0.9985	0.9952	0.9893	0.9804	0.9684	0.9532	0.9348	0.9134
3	1.0000	0.9999	0.9997	0.9990	0.9978	0.9957	0.9925	0.9880	0.9820
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9991	0.9984	0.9973
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
1	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
2	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
3	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
4	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
5	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
6	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128

7	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
8	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
9	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0079	0.0028	0.0008	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0356	0.0153	0.0056	0.0017	0.0004	0.0001	0.0000	0.0000	0.0000
4	0.1117	0.0573	0.0255	0.0095	0.0028	0.0006	0.0001	0.0000	0.0000
5	0.2607	0.1582	0.0846	0.0386	0.0143	0.0039	0.0007	0.0001	0.0000
6	0.4731	0.3348	0.2127	0.1178	0.0544	0.0194	0.0046	0.0005	0.0003
7	0.6956	0.5618	0.4167	0.2763	0.1576	0.0726	0.0239	0.0043	0.0027
8	0.8655	0.7747	0.6533	0.5075	0.3512	0.2054	0.0922	0.0256	0.0180
9	0.9579	0.9166	0.8487	0.7472	0.6093	0.4417	0.2642	0.1109	0.0866
10	0.9917	0.9804	0.9576	0.9150	0.8416	0.7251	0.5565	0.3410	0.2948
11	0.9992	0.9978	0.9943	0.9862	0.9683	0.9313	0.8578	0.7176	0.6775
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0016	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
8	0.0120	0.0075	0.0043	0.0022	0.0010	0.0003	0.0001	0.0000	0.0000
9	0.0652	0.0468	0.0316	0.0196	0.0107	0.0048	0.0015	0.0002	0.0000
10	0.2487	0.2033	0.1595	0.1184	0.0809	0.0486	0.0231	0.0062	0.0000
11	0.6323	0.5814	0.5241	0.4596	0.3873	0.3062	0.2153	0.1136	0.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=13

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.8775	0.7690	0.6730	0.5882	0.5133	0.4474	0.3893	0.3383	0.2935
1	0.9928	0.9730	0.9436	0.9068	0.8646	0.8186	0.7702	0.7206	0.6707
2	0.9997	0.9980	0.9938	0.9865	0.9755	0.9608	0.9422	0.9201	0.8946
3	1.0000	0.9999	0.9995	0.9986	0.9969	0.9940	0.9897	0.9837	0.9758
4	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9987	0.9976	0.9959
5	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9995
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
1	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017
2	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112
3	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461
4	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334
5	0.9991	0.9925	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905
6	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000
7	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095
8	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666
9	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9959	0.9888
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0041	0.0013	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0203	0.0078	0.0025	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0698	0.0321	0.0126	0.0040	0.0010	0.0002	0.0000	0.0000	0.0000

5	0.1788	0.0977	0.0462	0.0182	0.0056	0.0012	0.0002	0.0000	0.0000
6	0.3563	0.2288	0.1295	0.0624	0.0243	0.0070	0.0013	0.0001	0.0001
7	0.5732	0.4256	0.2841	0.1654	0.0802	0.0300	0.0075	0.0009	0.0005
8	0.7721	0.6470	0.4995	0.3457	0.2060	0.0991	0.0342	0.0065	0.0041
9	0.9071	0.8314	0.7217	0.5794	0.4157	0.2527	0.1180	0.0342	0.0242
10	0.9731	0.9421	0.8868	0.7975	0.6674	0.4983	0.3080	0.1339	0.1054
11	0.9951	0.9874	0.9704	0.9363	0.8733	0.7664	0.6017	0.3787	0.3293
12	0.9996	0.9987	0.9963	0.9903	0.9762	0.9450	0.8791	0.7458	0.7065
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0024	0.0013	0.0007	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000
9	0.0163	0.0103	0.0060	0.0031	0.0014	0.0005	0.0001	0.0000	0.0000
10	0.0799	0.0578	0.0392	0.0245	0.0135	0.0062	0.0020	0.0003	0.0000
11	0.2794	0.2298	0.1814	0.1354	0.0932	0.0564	0.0270	0.0072	0.0000
12	0.6617	0.6107	0.5526	0.4867	0.4118	0.3270	0.2310	0.1225	0.0000
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=14

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.8687	0.7536	0.6528	0.5647	0.4877	0.4205	0.3620	0.3112	0.2670
1	0.9916	0.9690	0.9355	0.8941	0.8470	0.7963	0.7436	0.6900	0.6368
2	0.9997	0.9975	0.9923	0.9833	0.9699	0.9522	0.9302	0.9042	0.8745
3	1.0000	0.9999	0.9994	0.9981	0.9958	0.9920	0.9864	0.9786	0.9685
4	1.0000	1.0000	1.0000	0.9998	0.9996	0.9990	0.9980	0.9965	0.9941
5	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
1	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
2	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
3	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
4	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
5	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120

6	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
7	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
8	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
9	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
10	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0022	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0114	0.0039	0.0011	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0426	0.0175	0.0060	0.0017	0.0003	0.0000	0.0000	0.0000	0.0000
5	0.1189	0.0583	0.0243	0.0083	0.0022	0.0004	0.0000	0.0000	0.0000
6	0.2586	0.1501	0.0753	0.0315	0.0103	0.0024	0.0003	0.0000	0.0000
7	0.4539	0.3075	0.1836	0.0933	0.0383	0.0116	0.0022	0.0002	0.0001
8	0.6627	0.5141	0.3595	0.2195	0.1117	0.0439	0.0115	0.0015	0.0008
9	0.8328	0.7207	0.5773	0.4158	0.2585	0.1298	0.0467	0.0092	0.0059
10	0.9368	0.8757	0.7795	0.6448	0.4787	0.3018	0.1465	0.0441	0.0315
11	0.9830	0.9602	0.9161	0.8392	0.7189	0.5519	0.3521	0.1584	0.1255
12	0.9971	0.9919	0.9795	0.9525	0.8990	0.8021	0.6433	0.4154	0.3632
13	0.9998	0.9992	0.9976	0.9932	0.9822	0.9560	0.8972	0.7712	0.7330
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0035	0.0020	0.0010	0.0004	0.0002	0.0000	0.0000	0.0000	0.0000
10	0.0214	0.0136	0.0080	0.0042	0.0019	0.0006	0.0001	0.0000	0.0000
11	0.0958	0.0698	0.0478	0.0301	0.0167	0.0077	0.0025	0.0003	0.0000
12	0.3100	0.2564	0.2037	0.1530	0.1059	0.0645	0.0310	0.0084	0.0000
13	0.6888	0.6380	0.5795	0.5123	0.4353	0.3472	0.2464	0.1313	0.0000
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=15

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.8601	0.7386	0.6333	0.5421	0.4633	0.3953	0.3367	0.2863	0.2430
1	0.9904	0.9647	0.9270	0.8809	0.8290	0.7738	0.7168	0.6597	0.6035
2	0.9996	0.9970	0.9906	0.9797	0.9638	0.9429	0.9171	0.8870	0.8531
3	1.0000	0.9998	0.9992	0.9976	0.9945	0.9896	0.9825	0.9727	0.9601
4	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9972	0.9950	0.9918

5	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
1	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
2	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
3	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
4	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
5	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
6	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
7	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
8	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
9	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
10	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963

13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0063	0.0019	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0255	0.0093	0.0028	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
5	0.0769	0.0338	0.0124	0.0037	0.0008	0.0001	0.0000	0.0000	0.0000
6	0.1818	0.0950	0.0422	0.0152	0.0042	0.0008	0.0001	0.0000	0.0000
7	0.3465	0.2131	0.1132	0.0500	0.0173	0.0042	0.0006	0.0000	0.0000
8	0.5478	0.3902	0.2452	0.1311	0.0566	0.0181	0.0036	0.0003	0.0002
9	0.7392	0.5968	0.4357	0.2784	0.1484	0.0611	0.0168	0.0022	0.0013
10	0.8796	0.7827	0.6481	0.4845	0.3135	0.1642	0.0617	0.0127	0.0082
11	0.9576	0.9095	0.8273	0.7031	0.5387	0.3518	0.1773	0.0556	0.0399
12	0.9893	0.9729	0.9383	0.8732	0.7639	0.6020	0.3958	0.1841	0.1469
13	0.9983	0.9948	0.9858	0.9647	0.9198	0.8329	0.6814	0.4510	0.3965
14	0.9999	0.9995	0.9984	0.9953	0.9866	0.9648	0.9126	0.7941	0.7570
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.92	p = 0.93	p = 0.94	p = 0.95	p = 0.96	p = 0.97	p = 0.98	p = 0.99	p = 1.00
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0007	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.0050	0.0028	0.0014	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000
11	0.0273	0.0175	0.0104	0.0055	0.0024	0.0008	0.0002	0.0000	0.0000
12	0.1130	0.0829	0.0571	0.0362	0.0203	0.0094	0.0030	0.0004	0.0000
13	0.3403	0.2832	0.2262	0.1710	0.1191	0.0730	0.0353	0.0096	0.0000
14	0.7137	0.6633	0.6047	0.5367	0.4579	0.3667	0.2614	0.1399	0.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Poisson Distribution Tables

x	μ									
	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.1135	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.1771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1734	.1781	.1823	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.0278	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001