INTRODUCTION TO STATISTICS MADE EASY

Tareq Alodat
UNSW Sydney

INTRODUCTION TO STATISTICS

MADE EASY

SECOND EDITION

Prof. Dr. Hamid Al-Oklah Dr. Said Titi Mr. Tareq Alodat

CONTENTS

The First Edition

The main objective of this book is to provide students of Preparatory Year Deanship, at King Saud University in Saudi Arabia, a textbook in statistics.

In fact, we found that most of the university statistics books are almost too much pages long, contain more material than can reasonably be covered in one term, and they are not readable for most of our students. In addition, part of the material covered in these books is redundant and unnecessary.

One of the most important features of this book is its readability, even for students who have weak reading statistical concepts. The book is written in a way that induces students to read and understand. The sentences used are short and clear. The concepts are presented in a student-friendly manner using an intuitive approach, and the examples are carefully chosen to reinforce understanding.

Each chapter opens with a list of objectives so that an instructor can tell at a glance what topics are covered in the chapter. In addition, students get an idea about the type of skills they should have acquired after studying the chapter.

In each section, there are a number of examples that have been worked out in a step-by-step detailed manner.

At the end of each section, there is a very carefully chosen set of exercises which cover all skills discussed in the section, and get the students thinking more deeply about the mathematics involved. . The exercises vary in difficulty and purpose. The reader will notice that the sets of exercises are much shorter than the traditional sets of exhaustive "drill and skill" questions, which build skill devoid of understanding. In other hand the first three chapters are supported by statistical package"SPSS" that helps the students to analyze and find the concepts in such chapters.

The topics in this book are organized in five chapters. In Chapter one, we introduce the basic concepts in statistics. Chapter two is devoted for organizing and graphing data set, and Chapter three is about numerical descriptive measure. Chapter four is concerned with basic concepts of probability and counting rule, and Chapter five is about random variables and their probability distributions.

Finally, we would like to thank his Excellency **Dr. Nami Al Juhani**, Dean of Preparatory Year at King Saud University, **Dr. Abdulmajeed Al-Jeriwi**, Vice Dean for Academic Affairs, **Dr. Obaid Al Qahtani** Chair of Basic Science Department for their continued support and encouragement. Many thanks go to all other faculty members, who helped us during the writing process; their input made this a much better book. Last, but definitely not least, we would like to thank our families for understanding and patience while writing this book.

Finally, it is pleasure to thanks and regards the co-authors of this book for their non-stopped efforts to reach the point of publishing the book. So, many thanks and respects to **Prof. Mohammed Subhi Abu-Saleh, Dr. Sharhabeel Alaidi and Mrs. Dareen Omari** for their unlimited useful feedback and for the remarkable touches. We also take this opportunity to express a warm gratitude to **Mr. Anas S. Alakhras and Mr. Fadi Hassan** for their help in publishing this book.

The Authors

March 10, 2014

The Second Edition

Based on the directives of the Department of Basic Sciences at the Deanship of the preparatory year in to develop some of the decisions of the books studied by the students of the preparatory year at King Saud University, it has included the development of the book Introduction to Statistics for authors Dr. Said Titi, Khaled Khashan and Mr. Tareq Alodat.

To this aim, it formed a committee composed of Prof. Dr. Hamid Owaid Al-Oklah, Dr. Said Titi and Mr. Tareq Alodat , so that the committee review the contants of the book and is interested in developing. In addition to that the design process is Assigned by Alodat.

The book has been reviewing in question above by Prof. Dr. Al-Oklah and author Dr. Titi where corrected most of the typos that have been monitored, and corrected some of the concepts, and then added several paragraphs, examples and exercises necessary for the book, and delete others for lack of necessity, and moreover has been included Dictionary (English - Arabic) for most of the scientific terms contained in this book.

At last we draw thanks to the Deanship of the preparatory year and also to the presidency of the Department of Basic Sciences specially for **Dr. Obaid Al Qahtani** Chair of Basic Science Department for their trust in us in the development of this book.

We hope from God that we have been successful in our work and God Crown success.

Prof. Dr. Hamid Al-Oklah

Dr. Said Titi

June 22, 2015

PREFACE

Statistics book covers many statistical fields. Our life is full of events and phenomena that enhance us to study either natural or artificial phenomena could be studied using different fields of science like physics, chemistry, and mathematics. The goal of this book is to connect those concepts with the advanced statistical problems.

Statistics is used in a variety fields like business and engineering and science. We can sea there are many applications of statistics in those fields, the applications of statistics are many and varied; people encounter them in everyday life, such as in reading newspapers or magazines, listening to the radio, or watching television. Since statistics is used in almost every field of endeavor, the educated individual should be knowledgeable about the vocabulary, concepts, and procedures of statistics.

CHAPTER

1

Basic Concepts in Statistics

OBJECTIVES

1. Understand the role of statistics in real life.
2. Understand the definition of basic Statistical concepts.
3. Distinguish between descriptive and inferential statistics.
4. Obtain types of Variables.
5. Determine the measurement level for each variable.
6. Obtain the basic sample techniques.
7. Explain the difference between an observational and experimental study.
8. Take an overview about applications using statistical software (SPSS).

1.1 Statistical Concepts

Our life is full of events and phenomena that enhance us to study either natural or artificial phenomena could be studied using different fields one of them is statistics. For example, the applications of statistics are many and varied as follows:

-People encounter them in everyday life

-Reading newspapers or magazines,

-Listening to the radio, or watching television.

Since statistics is used in almost every field of endeavor, the educated individual should be knowledgeable about the vocabulary, concepts, and procedures of statistics.

Definition 1.1.1

Statistics is a branch of science dealing with collecting, organizing, summarizing, analysing and making decisions from data.

Statistics is divided into two main areas, which are **descriptive** and **inferential** statistics.

A Descriptive Statistics

Suppose that a test in statistics course is given to a class at KSU and the test scores for all students are collected, then the test scores for the students are called data set (the definition of this term will be discussed deeper in section 1.2). Usually the data set is very large in the original form and it is not easy to use it to draw a conclusions or to make decisions while it is very easy to draw conclusions from summary tables and diagrams than from such original data. So reducing the data set to form more control by constructing tables, drawing graphs and provide some numerical characteristics for which is a simple definition to introduce descriptive statistics.

Definition 1.1.2

Descriptive statistics deals with methods for collecting, organizing, and describing data by using tables, graphs, and summary measures.

B Inferential Statistics

The set of all elements (observations) of interest in a study is called a population, and the selected numbers of elements from the population is called a sample. In statistical problems we may interest to make a decision and prediction about a population by using results that obtained from selected samples, for instance we may interest to find the number of absent students at PY on a certain day of a week, to do so, we may select 200 classes from PY and register the number of students that absent on that day, then you can use this information to make a decision. The area of statistics that interest on such decision is referred to inferential statistics.

Definition 1.1.3

Inferential statistics deals with methods that use sample results, to help in estimation or make decisions about the population.

During this section, we will clarify the meaning of population, sample, and data. Therefore, the understanding of such terms and the difference between them is very important in learning statistics. For example, if we interest to know the average weights of women visited diet section in a hospital during specified period of time, then all women who visited that section represents the study population.

Definition 1.1.4

A population is the set of all elements (observations), items, or objects that bring them a common recipe and at least one that will be studied their properties for a particular goal. The components of the population are called individuals or elements.

Remark

Note that a population can be a collection of any things, like Ipad set, Books, animals or inanimate, therefore it does not necessary deal with people.

Any collection of things, including a joint gathering recipe at least one to be examined for a particular purpose, called a statistically population (or population as a matter of shortcut). The components of the population are called individuals or elements.

Example 1

a. In a study of the average number of students in secondary schools in Riyadh city, where there are different stages of the students, such as first, second and third secondary, as well as there are male and female, but they all gathered, including prescription study in high school. Therefore, we find that high school students in Riyadh make up a population.

b. In a study of the evolving condition of the patients in a hospital, where there are many people of different types of diseases, but they all bind them recipe disease, so patients that in the hospital make up a population.

c. In a study to determine the technical condition of the aircraft of the Gulf Cooperation Council (GCC), where travel aircraft, military training aircraft, and Helicopters ...., but they are all characterized by their ability to fly, so the aircraft in the Gulf Cooperation Council (GCC) are population

Note that a population can be a collection of any things, like set of trees, people, animals or inanimate (books, cars, metal...). Therefore it does not necessary deal with a people.

Definition 1.1.5

A sample is a subset of the population selected for study.

Referring to the example of interest to know the average weight of women that visited diet section, in this case the registered weights of some women represent a sample.
In practical life there are many ways to get a sample from the population under study, for example; face-to-face interview, online electronic questionnaires, paper questionnaires and using telephones.

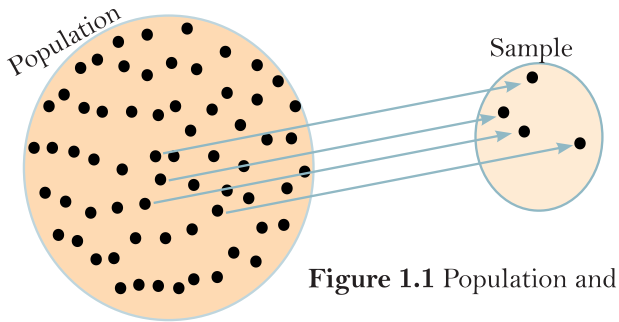


Figure 1.1 Population and Sample

Let us discuss an example on determining the population and the sample for a study

Example 2

If we take a class of the students in Stat 140 course at PY, then: all the students registered in the course represent the population, and any class of them is represent a sample.

1.2 Variables and Types of Data

Basic terms that will be used frequently in this section, and they are very important tools in statistical problems, such terms are, an **element**, a **variable** and their types, a **measurement**, and a **data set,** Therefore to understand such terms, it is necessary to illustrate the following definitions.

Remark

Any study is based on a problem or phenomenon such as heavy traffics, accidents, rating scales and grades or others. The researcher should define the variables of interest before collecting data.

Definition 1.2.1

An element (or member of a sample or population) is a specific subject or object about which the information is collected.

Example 1 below discuss the definition of an element numerically

Example 1

The following table gives the number of snake bites reported in a hospital in 3 cities (A, B, C).

City	Number of Snake Bites
A	10
B	17
C	11

Each one of the cities is a member, that is; city A is a member, city B is a member, and also city C is a member.

Definition 1.2.2

A variable is a characteristic under study that takes different values for different elements.

For example, if we collect information about income of households, then income is a variable .These households are expected to have different incomes; also, some of them may have the same income. Note that a variable is often denoted by a capital letter like $X, Y, Z, \ldots$ and their values denoted by small letters for example $x, y, z, \ldots$.

Definition 1.2.3

The value of a variable for an element is called an observation or measurement.

The following is an example to explain the difference in the meaning between variable and the measurement.

Example 2

Referring to example 1, we see that the variable (let X for example) is the number of snake bites and each one of the number of bites 10, 17, 11 represents an observation or measurement. Where we have $X(A)=10$, $X(B)=17$ and $X(C)=11$.

We know that the variable is a characteristic under study that takes different values for different elements. In statistics, we have two types of variables according to their elements; first type is called **quantitative** variable and the second one is called **qualitative** variable.

When a subject can be measured numerically such as (the price of a shirt), then the subject in this case is quantitative variable. The following definition provides us with this concept.

Definition 1.2.4

Quantitative variable gives us numbers representing counts or measurements.

When a subject cannot be measured numerically such as (eye color), then the subject in this case is qualitative variable. The following definition provides us with this concept.

Definition 1.2.5

Qualitative variable (or categorical data) gives us names or labels that are not numbers representing the observations.

Remark

Quantitative variables give us quantitative data and inquires about the phrase "how much", while the qualitative variables give us the qualitative data and inquires about the phrase "what or what is".

The following examples illustrates the two type of variables

Example 3

The following table shows some examples of the two types of variables

Quantitative variable gives us quantitative data	Qualitative variable gives us qualitative data
The age of people in years 19, 2, 45, 23, 88, ...	The gender of Organisms Male, Female ......
Number of children in family 5, 2, 4, 1, 14, ...	Results tossed a coin twice $HH,\ HT,\ TH,\ TT$ (H=Head, T=Tail)
The heights of buildings in meters 15, 5.6, 12.7, 105, 27, ...	Eye color of people Black, Brown, Blue, Green, ...
The weights of cars in tons (ton=1000 Kg) 2.35, 1.65, 2.05, 2.10, 1.30, ...	Religious affiliation Muslim, Christian, Jew, ...
The speed of a car going on a main road in Km 110, 105, 85, 120, 90, ...	The pressure in a boiler High, Moderate, Low

Moreover, the variables measured in quantitative data divided into two main types, **discrete** and **continuous**. A variable that assumes countable values is refer to discrete variable, otherwise the variable is a continuous one. Accordingly, we provide the following definitions.

Definition 1.2.6

Discrete variables assume values that can be counted.

In following we illustrate some examples on a discrete variable

Example 4

-The number of children in a family, , where we have 1,2,3, ... or k children.

-The number of students in a classroom, where we have 21, 25,32,18 and so on.

-Number of accidents in a city, where we have 1,2,3,... or k accidents.

The other type of quantitative variable is the continuous variable which is assumed uncountable values, and offer us the following definition.

Definition 1.2.7

Continuous variables assume all values between any two specific values, i.e. they take all values in an interval. They often include fractions and decimals.

In the following we illustrate some examples on a continuous variable

Example 5

- Temperature: For example the temperature in Riyadh city in last summer was between 15 and 56, i.e. the temperature $t \in [15, 56]$.

- Age: For example the age of a horse is between 0 (Stillborn) and 62 years (Said the oldest horse was 62 years, but the middle age of a horse is 30 years), i.e. the age of a horse $x \in [0, 62]$

- Height: For example the height of a student in a Country is between 110 cm (person elf) and 226 cm (person giant), i.e. the height of a student $x \in [110, 226]$

Variables classified according to how they are categorized or measured. For example, the data could be organized into specific categories, such as major field (mathematics, computers, etc.), nationality or religious affiliation. On the other hand, can the data values could be ranked, such as grade $(A,\ B,\ C,\ D,\ F)$ or rating scale (poor, good, excellent), or they can be classified according to the values obtained from measurement, such as temperature, heights or IQ scores. Therefore we need to distinguish between them through the measurement scale used. There are four levels of measurement **scales**; **nominal**, **ordinal**, **interval**, and the **ratio** level of measurement, the difference between these four levels is explained in the following definitions.

Definition 1.2.8

The nominal level of measurement classifies data into mutually exclusive (disjoint) categories in which no order or ranking can be imposed on the data.

The following examples include nominal level of measurements in different cases.

Example 6

- Gender: Male, Female.
- Eye color: Black, Brown, Blue, Green, ...
- Religious affiliation: Muslim, Christian, Jew, ...
- Nationality: Saudi, Syrian, Jordanian, Egyptian, Pakistani, ...
- Scientific major field: statistics, mathematics, computers, Geography, ...

When the classification takes ranks into consideration, the ordinal level of measurement is preferred to be used. The following definition provided us this concept.

Definition 1.2.9

The ordinal level of measurement classifies data into categories that can be ordered, however precise differences between the ranks do not exist.

The following examples include some ordinal level of measurements.

Example 7

Grade $(A,\ B,\ C,\ D,\ F)$: Grading technique is the most common example on ordinal level. For example we find that the system of appreciation in Saudi universities are (in descending order) $A^{+},\ A,\ B^{+},\ B,\ C^{+},\ C,\ D^{+},\ D,\ F.$

- Rating scale (bad, good, excellent and so on ...): To test the quality of the canned product, we find that the state of the tested object either excellent or good or bad.

- Ranking of football players: A football player can be ranked in first grade, second grade, third grade, ...

- Ranks of university faculty members: Academic ranks usually classified as professor, associate professor, assistant professor , and instructor.

The third level of measurement is called interval level. The following definition provided us this concept.

Definition 1.2.10

The interval level of measurement orders data with precise differences between units of measure. (in this case there is no meaningful zero). On the other hand, the resulting measurement values belong to an interval of the real numbers.

Example 8

- **IELTS:** An International English Language Testing System. ILETS is an international system to test the English language in order to study and work. The degree x to which the grant will be between zero and 9, i.e. $x \in [0, 9]$
- **TOEFL:** Test of English as a Foreign Language. TOEFL is a standardized test of English language proficiency for non-native English language speakers wishing to enroll in some universities in the world.
- **SAT** score and IQ test: The SAT is a standardized test widely used for college admissions in some universities in the world. It is a good predictor of a student's performance in the first year of college. Degree to which the grant will be between zero and 2400, i.e. $x \in [0, 2400]$
- **Temperature:** When the degrees of temperatures are measured in Celsius or Fahrenheit, then the values that we obtain from absolute zero (-273.15 but without this degree) extends to millions as is the case in the sun and stars.

Note that if we compare between two temperature degrees, like 30°C and 60°C we can't say that 60°C is as high as twice the degree 30°C; but we can say there is a 30°C difference between them. In the sense that we can not be compared to some of the quantities to others in this case. On the other hand, if the temperature of something equals to zero, that does not mean it does not have a temperature.

That mean in the Celsius temperature the zero means there is a temperature and it is very cold that is the zero does not mean nothingness. The data at this level do not have a natural zero starting point. The measurements that rely (or that adopt) zero as starting point called ratio level and offered us the following definition

Definition 1.2.11

The ratio level of measurement is the interval level with additional property that there is also a natural zero starting point. In this type of measurement zero means nothingness. Another difference lies in that we can attribute some of the quantities to others.

The following examples include some of ratio level measurement, we note what is the meaning of something that takes a temperature equal to zero

Example 9

- Distance: The distance between two cities X and Y, where we find that the measurement is an interval level, but because of that we can say that the distance between the two cities X and Y is equal twice the distance between the cities X and Z, they become standard ratio scale. Note that here zero has a meaning, because if the distance is equal to zero, it means that the city (position) itself. Here we note that the concepts of length and height are a special case of the distance concept.
- Age: The ages of people fall under this category of measurements, because the zero here means that the person was born dead and that old equals zero.

- Time: The time required to get from home to work is a measurement of type ratio level and zero here means that it has not yet kicks off. Note that here we can attribute some of the time to others, if we say the time required to get from home to work is equal twice the time needed to get home from work.

- Salary: The value of salary for someone is a measurement of type ratio level, where we can attribute values of wages to each other, as if to say that the person X receives a salary twice the salary of the person Y. And zero here means that the person did not receive a salary.

- Weights: If we take weights of fruit boxes, then note that the values are of measurements of type ratio level, due to the possibility of the weights attributed to each other, where we can say that an apple box weight is equal twice the weight of the orange box, and the zero here it is linked to accurately device which measures the weight. And so that there is nothing

on the earth has no weight because of Earth's gravity. For example, if we have the weight of a box containing apple and using the balance of accuracy 200 grams, we will not get to zero absolute (it is enough having one apple in the box so that does not refer to zero), while if we weigh it in another balance accuracy tons, the balance will refers to the value zero even if the box full of apples.
Here there is a meaningful zero.

The graph below summarize the classification of variables

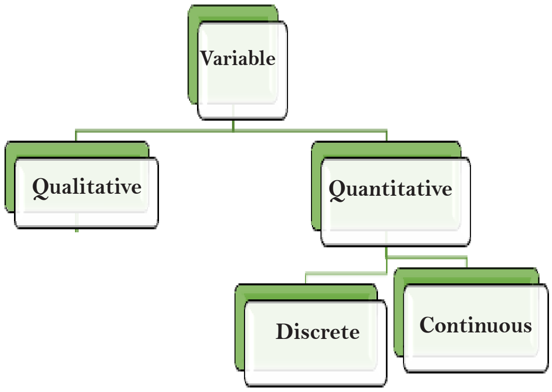


Figure 1.2: Classification of variables

1.3 Sampling Techniques

Its known that in some cases, it's hard to study a large population in order to make conclusions about certain phenomena, for example if we interested in studying the obesity in kingdom of Saudi Arabia, imagine that the researcher has a limited period of time say three months, its ' impossible' to survey all citizens in KSA to make conclusions about such phenomenan during the determined period of time, so that sampling methodology is the best solution in order to perform the study and get representative results during shorter period and also it saves efforts and money.

Sampling methodology as illustrated in section (1.1) suggests selecting a portion of elements of population under study in order to make statistical analyses to make decisions about phenomenan under study. Therefore, sampling is not include the selection of elements arbitrary. So that, there are several techniques of sampling were established according to the type of analysis used

Some of these techniques are the **simple random sampling** method, the **systematic** method, the **stratified** method and the **clustered sampling** method. Differences between such methods refer to many circumstances such as population size, degree of accuracy determined by the researcher, type of elements of population and the number of categories in the population under study.

Now, let us discuss the mentioned methods in details starting with the simple random sampling method.

A Simple Random Sampling Method

It's the simplest method for sampling and it is applicable when the population is slightly small. In order to get a sample of this type the elements of population should be to achieve the following conditions:

1. All elements of population have the same chance of choice,
2. All elements of population are independent.

After verification of the fulfillment of these conditions the elements of population take serial numbers and then use one of the methods that used in randomization order to pull the required elements of the sample, and in this regard can be used a table of random numbers (See table 1, bellow).

92	63	07	82	40	19	26	79	54	57
79	44	57	87	35	71	54	94	48	43
59	65	47	19	66	27	38	65	00	04
31	52	44	95	87	76	61	23	97	89
06	34	87	69	38	90	37	95	13	92
28	70	35	17	09	94	45	64	83	96
68	10	88	92	66	94	73	09	57	61
99	93	89	07	04	93	62	16	63	30
91	54	37	31	96	34	44	96	35	13
42	10	30	27	81	73	92	05	62	97
17	13	82	75	42	52	24	21	61	18
28	29	71	42	80	54	52	42	16	18
09	33	15	67	12	51	33	30	62	87
31	29	50	42	04	93	71	25	12	87
36	14	61	55	60	27	59	24	20	89
29	55	31	84	32	13	63	00	55	29
23	50	12	26	42	63	08	10	81	91
57	88	88	58	46	67	96	70	78	35
55	33	67	12	64	88	47	20	43	34
10	08	71	00	72	55	98	06	46	88

Table 1: Random Numbers

Three steps to use the random number table such steps are:
1. Close your eyes
2. Point your finger anywhere in the random numbers table.
3. Open your eyes and begin reading the digits beginning where your finger touches the table)

Example 1

To select a random sample consisting of 10 elements out of 90 elements, it is necessary to number each element from 01, 02, 03, $\cdots$ to 90. Then select a starting number by closing your eyes and placing your finger on a number in the table 1. Suppose, in this case your finger is landed on the number 19 in the forth column. Then proceed downward until you have selected 10 different numbers between 01 and 90. When you reach the bottom of the column, go to the next column. If you select a number greater than 90 or the number 00 or duplicate number, just omit it. In our example, we will use the elements numbered: 19, 69, 17, 07, 31, 27, 75, 42, 67, and 55.

Even that simple sample method is easy to perform, but it has some disadvantages that make it not the best choice to use, especially when we are talking about large populations, it costs more money and much time, many samples can be selected using this method, but they might give same results. So that, statisticians and researchers developed other alternatives to be used to get more convenient.

B Systematic Sampling Method

Suppose we want to take a sample with size n using this method, we are including the following:

1- We giving the elements of the population serial numbers from 1 up to N

2- Determining an interval (called the withdrawal period). This interval can be computed their width by dividing the size of the population that we are interested by the required sample size.

$$k = \frac{N}{n}$$

3- Then we randomly select number located between 1 and k (Let s, for example), so the element that holds this number s is the start element in the sample.

4- Take elements from population that bear numbers $s + t \cdot k$ with $1 \le t \le n - 1$. So we get the required sample.

To understand this, assume example 2

Example 2

Suppose there are 1000 elements in the population, and a systematic sample of 40 elements is needed, then

1- We giving the elements of the population serial numbers from 1 up to 1000

2- Determining an interval (called the withdrawal period) by the following relation:

$$k = \frac{N}{n} = \frac{1000}{40} = 25$$

3- Then we randomly select number located between 1 and k (Let s =19, for example), i.e. the start element is the 19^{th} in the population.

4- We take elements from population that bear numbers $s + t \cdot k$ with $1 \le t \le n - 1$, i.e. the elements:

19, 44, 69, 94, 119, 144, 169, 194, 219, 244 , 269, 294, 319, 344, 369, 394, 419, 444, 469, 494,519, 544, 569, 594, 619, 644, 669, 694, 719, 744, 769, 794, 819, 844, 869, 894, 919, 944, 969, 994.

So we get the required sample.

The preceding two methods discussed the sampling techniques under population without subgroups. The situation is different when the population is composed of several subgroups, so that a technique is developed called stratified sampling technique.

C Stratified Sampling Method

In statistics, a subset of a population share some characteristics is called a 'stratum' the plural is strata. In such condition, the stratified sampling method is used and these subsets are selected randomly.

To explain this consider the following example

Example 3

Suppose that PY administration want to measure the level of satisfaction of students about certain issue and whether there are differences between the opinions of the scientific path students and the humanities path students. So that, the administration will select students from each group to use the sample, its reasonable that, the size of each sample is proportional to the size of its relative to the whole population.

Example 4

If you have 3 strata with 100, 200 and 300 population sizes respectively. And the researcher chose a sampling fraction of $\frac{1}{2}$. Then, the researcher must randomly sample 50, 100 and 150 subjects from each stratum respectively.

The fourth technique that can be used for sampling is the clustered sampling technique.

D Cluster Sampling Method

The difference between the stratified sampling method and the clustered is that, in case of stratified the researcher select a random sample of elements from population strata and the analyses are performed on the elements directly, while in other case; the analyses are performed on the clusters chosen randomly from the population. Usually, each cluster consists of heterogeneous elements based on geographical bases. The advantage of this method is that it's cheaper than other methods.

Cluster sampling is used when the population is large or when it involves elements residing in a large geographic area. To understand this method, let use consider an example

Remark

Clustered sampling technique might be conducted on more than one stage.

Example 5

If a researcher interested in surveying the number of students in Government of KSA universities who own American made cars. Assume that there are 25 universities in KSA. To do so, the researcher can select 5 universities and survey all students in these universities using cluster sampling method. This type of clustered sampling is called 'single- stage-cluster sampling'.

Example 6

Referring to example 5, suppose that the researcher is interested in knowing the specialization of students who own American made cars in KSA governmental of universities. Assume that the needed sample size is 3000 students.First stage the researcher will chose 5 universities, and then the determined sample size is 3000 students. These can be selected from the five universities using the simple random sampling method or using the systematic method. In this case, we have two stage cluster sampling.

As mentioned before, the type of sampling method differs depending on the sample size and the acceptable degree of accuracy. After discussing sampling techniques in section 1.3, students should know how to complete the process of performing a statistical study using a scientific methodology. Section 1.4 discusses the main concepts of experiments and the difference between types of studies.

1.4 Observational and Experimental Studies

Any researcher wishes to study a certain phenomena has to go through an organized mechanism in order to get reasonable conclusions, starting from determining the main objective of the study till the stage of writing conclusions and recommendation . Many of performed studies involve an experiment to test a proposed claim by the researcher, while other studies do not need to involve such experiments because the researcher can collect information using past observations.

Statistical studies are classified into two types of studies according to how researcher gets observations. In an observational study, the researcher is interested in studying something happened in the past or happening at the moment of performing the study, then (he/she) makes statistical analysis on collected observations to take the right decision. In other words, it's used when the researcher interested in studying the correlation between two or more variables, for example number of studying hours and GPA.

A survey is a kind of an observational study which can be performed by several ways. In experimental studies, the researcher is interested in studying elements of population after distribute them into groups and each group is being studied using different treatments. That is, the researcher applies factor(s) on the groups under study in order to compare them. This type of studies also concerns about the effect of a variable called independent variable on other variable(s) called dependent variable(s). One more thing is that experimental studies can be used to study phenomena include human intervention, but the situation is different in case of observational studies, clinical trials are good example on experimental studies.

Examples 1 and 2 explain the difference between the two types of studies.

Example 1

Assume that a researcher wishes to know if there is a relationship between the number of absence hours for KSU.PY graduated students and their GPAs, to do this the researcher determined a sample size of 3000 students were surveyed in order to analyze the data collected and conclusions were made. Since the researcher wanted to compare between students according to their absence hours in order to study its effect on student's GPA, it represents the process of studying something happened in the past and no treatments were used.

Note that you can observe and measure, but not modify the study. In this study, the researcher manipulates one of the variables and tries to determine how the manipulation influences other variable. So, the variable which is manipulated by researcher is called the independent variable or explanatory variable, and the outcome variable is called the dependent variable.

Example 2

Suppose that a doctor in a hospital is interested in studying the effect of a new medicine on his patients who suffered from cancer. He selected 20 patients and divided them into 2 groups A and B, he applied the new medicine for group A, and group B still take the old one. After one year, he studied if the new medicine has good effects on their health status using statistical analyses.

It's clear from example 2 that the researcher used experimental study.

At the end of chapter 1, students should realize how to perform a statistical study using scientific methodology.

Next chapter discusses the process of manipulating data and different statistical techniques to analyze collected data.

Exercises On Chapter One

1 Give an example for each of the following concepts:

(i) Discrete variable.
(ii) Continuous variable.
(iii) Nominal-level measurement.
(iv) Ordinal-level measurement.
(v) Interval-level measurement.
(vi) Ratio-level measurement.

2 Classify each according to level of measurement with the interpretation of the meaning of zero if exist.

(i) Ages of students in the college (in years).
(ii) Ages premature babies in the maternity hospital (in hours).
(iii) Color of eyes of people.
(iv) Colors of Spectrum of light.
(v) Rankings of football players.
(vi) Temperatures inside room (in Celsius).
(vii) Temperatures inside high cold cooling device in a labor (in Kelvin).
(viii) Nationalities of the workers in Riyadh.
(ix) Salaries of employees in the college.
(x) Weights of boxes of fruits.
(xi) Criminal cases in court

3 Classify each variable as qualitative or quantitative.

(i) Time neededtofinish the exam.
(ii) Colors of basketball team T-shirts.
(iii) Weights of luggage of passengers.
(iv) Classification of children in a day care center according to gender.
(v) Marital status of faculty members in King Saud University.
(vi) Horsepower of tractor engines.

4 Classify each variable as discrete or continuous

(i) Lifetime (in hours) of table lamps.
(ii) Number of cars rented each week.
(iii) Number of cups sold each day by coffee shop.
(iv) Weights of boys in a school.
(v) Capacity (in gallons) of ten jugs of oil.

5 Classify each sample as simple random, systematic, stratified, or cluster.

(i) Out of every 50 cars manufactured is checked to determine its gear.

(ii) Out of every 10 customers entering a shopping mall is asked to select his favourite store.

(iii) Assistant professors are selected using random numbers to determine annual salaries.

6 For each of these statements, define the population and state how a sample might be obtained.

(i) Every 40 minutes, 1 people die in car crashes and 120 are injured is Saudi Arabia.

(ii) The average cost of an airline mail is 20SAR.

7 The table below shows the number of new AIDS cases in the U.S. in each of the years

Year	New AIDS cases
1989	33,643
1990	41,761
1991	43,771
1992	45,961
1993	103,463
1994	61,301

Classify the study as either descriptive or inferential..

8 The table below shows the average income by age group for the residents in Riyadh in the year 2012. The average incomes for each age group are estimates based on a sample of size 100 from each group.

Age group	Average income
18 - 24	33,643
25 - 39	41,761
40 - 54	43,771
55 - 70	45,961
Over 70	103,463

Classify the study as either descriptive or inferential.

9 The table below shows the total number of births in the KSA and the birth rate per 100 population in each of the years 2003-2007.

Year	Births	Birth rate
2003	473,725	24
2004	477,251	24
2005	482,252	23
2006	488,980	23
2007	495,063	22

Classify the study as either descriptive or inferential.

10 Based on a random sample of 100 people, a researcher obtained the following estimates of the percentage of people lacking health insurance in Dammam.

Age	Percentage not covered
18 - 24	28.2
25 - 39	24.9
40 - 54	19.1
55 - 65	16.5

Classify the study as either descriptive or inferential.

11 100,000 randomly selected adults were asked whether they drink at least 48 oz of water each day and only 45% said yes. Identify the sample and population.

12 The manager of a car dealership records the colors of automobiles on a used car lot. Identify the type of data collected.

13 A postal worker counts the number of complaint letters received by the KSA Postal Service in a given day. Identify the type of data collected.

14 An usher records the number of unoccupied seats in a movie theater during each viewing of a film. Identify the type of data collected.

15 If "color of a state" is defined as {red, blue, purple}, then the variable is ....

16 If ”how it votes” is defined as {very liberal, liberal, neutral, conservative, very conservative}, then the variable is ....

17 The researcher chose to measure age as a number 18 to 110. What level of measurement is age for this research question.

SPSS Statistical Applications

One of the mostly statistical packages used in data analysis is SPSS software, the abbreviation SPSS means Statistical Packages for Social Sciences. Throughout, this book we will introduce to you how to handle problems related to statistics using SPSS environment version 17.0, starting with data entry, filtering data, and ending by the interpretation of the results.

Moreover, you will find at the end of each section an application that explains how to solve statistical problems using SPSS, like construction of frequency table, graphing variables and computing statistical estimates and measurements, e.g. (mean, median, variance, etc ...).

Mainly, SPSS environment has two views or windows; data view and variable view. Figure 1 below shows how data view looks like.

Remark

Data view in SPSS resembles the view in Microsoft Excel software.

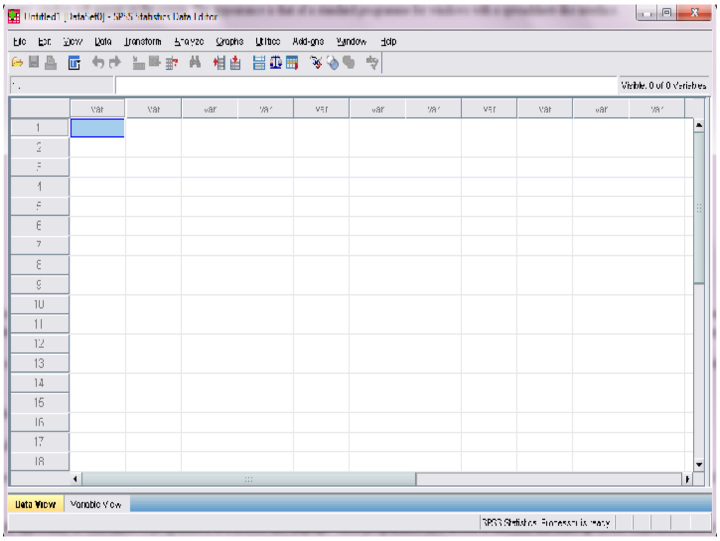


Figure1. Data view in SPSS environment

As shown in Fig. 1, data view appears when you click the SPSS icon, and it's used to input the data that we are interested in studying. As a beginner in SPSS, you will need to know four main menus; File, Data, Analyze and Graphs menus.

File menu contains commands that used to: create new file, open files (or saved database files on the form of excel, access and notepad) and to save the created SPSS file. See figure 2.

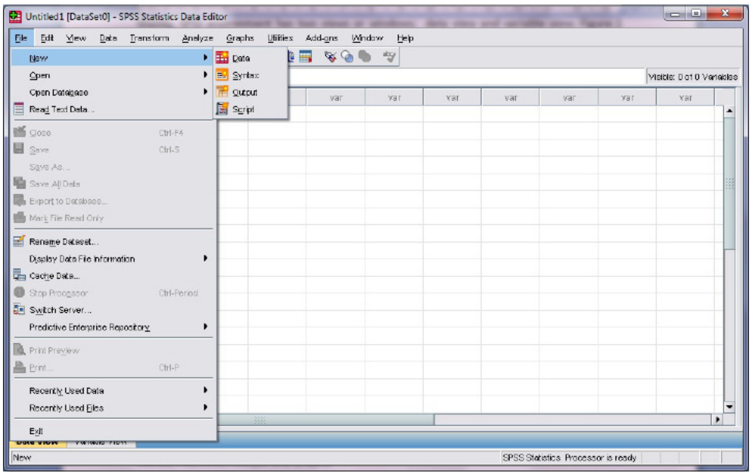


Figure2. File menu in SPSS environment

Data menu is used for manipulating data like defining, sorting, merging and selecting some cases of recorded data. Moreover, it can be used to combine data cases from two or more files

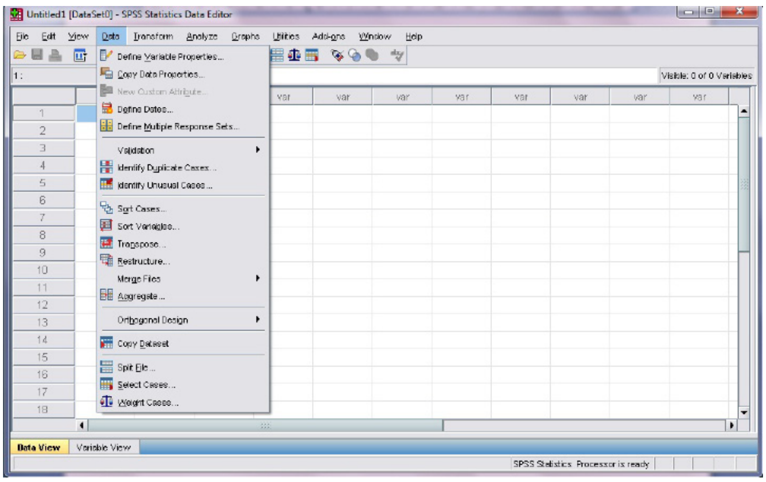


Figure3. Data menu in SPSS environment

Analyze menu contains the core options of SPSS, many commands for analyzing data can be found in this menu. Through this book, students need to focus on **descriptive statistics** submenu which used to make basic analyses as will be shown in coming sections.

Remark

Data management can be performed using data menu.

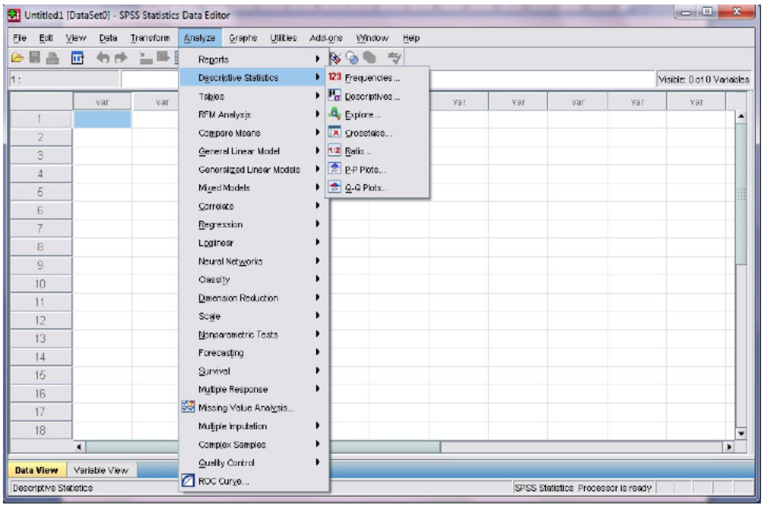


Figure 4. Analyze menu in SPSS environment

Remark

The shortest and meaningful name for variable is the preferred one.

And the forth menu that we are interested in is **Graphs menu**, and the usage of this menu can be known from its' name. Different types of graphs can be achieved according to the type of variable as you learned in section 1.2.

Major types of graphs are; bar chart, line, pie chart, boxplot, scatter/ Dot and Histogram. Some of these types are used to describe the shape of data like (Bar, Pie, etc..) and other used to study the dispersion of data like (line chart and scatter/ Dot chart) and finally graphs can be used in order to check the normality of collected data (i.e. whether it has a normal distribution or not), to do this, one can use histogram and other plots like P-P plot from analyze menu.

One more thing the student should know is: you can edit and change the format of the graph after plotting it Figure 5, below shows how graphs menu look like.

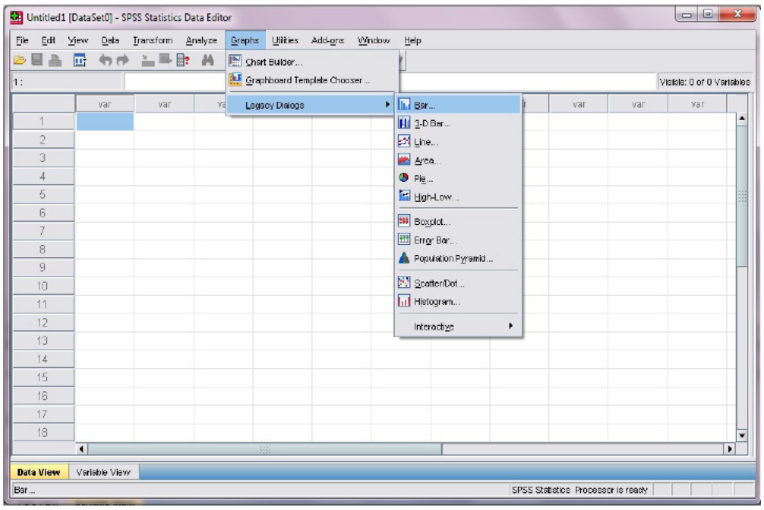


Figure5. Graphs menu in SPSS environment

All previous illustration can be considered as an introduction of data view, more details about how to use each menu will be explained in the next chapter .Now, the second main view (window) is the **variable view**.

Variable view contains the characteristics of each variable involved in analysis starting with: name of variable, type of variable, label, values, missing, align and measure as shown in fig. 6. **The name of a variable** must be meaningful, unique, not to exceed 40 letters, starts with a letter and do not use space or mathematical signs (- , + , * , /), but you can use underscore (_).

Type of a variable, as shown in figure 6 which displays how **variable view** looks like, and when the user press on cell of variable type a dialog box appears to select the type according the data understudy. Numeric type can be used for quantitative variables and the user can determine the number of decimal places for the numeric data.

Label column is used to clarify and give more details about the variable name, since the name should be simple and meaningful.

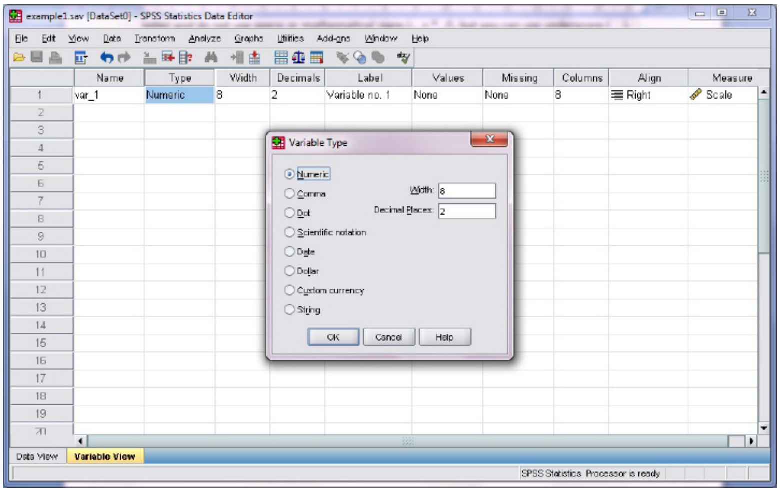


Figure6. Variable view in SPSS environment

The next column as shown in figure 6 is "**Values**" which allows the user to enter values for variables with more than one level, for example if we are in collecting data about gender which has two levels; male and female, one can give the value 1 for male and 2 for females, then press "add" each time in order to make data entry simple and to make numerical analyses about "gender" variable. See figure 7

Remark

Some version of SPSS do not allow to name file in Arabic language.

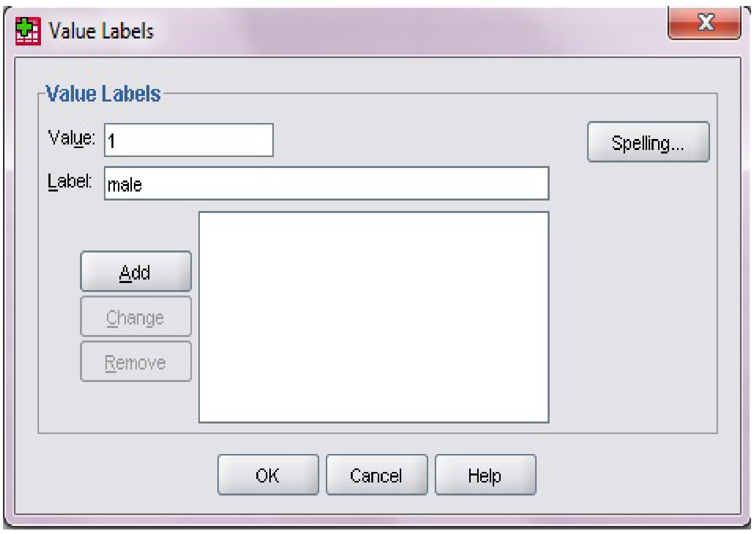


Figure7. Value label dialog box

Measure column is a drop list of three types of measurements; first one is **scale** and this type is assigned for quantitative variables, the other two types are **ordinal** and **nominal** areas signed to qualitative variables, but the difference is that the **nominal level** of measurement classified data into mutually exclusive (disjoint) categories in which no order or ranking can be imposed on the data, examples on nominal level:

- Gender (male, female)
- Eye color (blue, brown, green, hazel)
- Surgical outcome (dead, alive)
- Blood type (A, B, AB, O)

The **ordinal level** of measurement classified data into categories that can be ordered, however precise differences between the ranks do not exist, examples on ordinal level:

- Education level (elementary, secondary, college)
- Pain level (mild, moderate, severe)
- Agreement level (strongly disagree, disagree, neutral, agree, strongly agree)

Figure 8 displays the shape of measure column in the variable view.

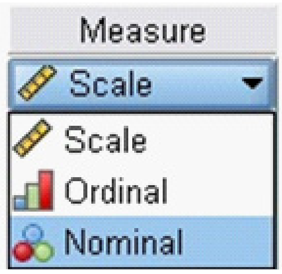


Figure8. Measure level drop list in variable view

Finally, when user want to save work on SPSS data file is saved with an extension (*.sav), and the output data (*.spv), it's preferred to save your file in English letters and meaningful. More applications on SPSS will be discovered in next chapters.

CHAPTER 2

Organizing and Graphing Data

OBJECTIVES

1. Define raw data.
2. Organize and graph Qualitative data.
3. Graph Qualitative data.
4. Organize and graph quantitative data.
5. Graph quantitative data.

In this chapter we will learn how to organize and display the Qualitative and Quantitative data.

2.1 Raw Data

In many practical problems, after collecting data, the set of all information that obtained from each element of a sample or population are recorded in a sequence in which it become available. The recorded sequence is random and it is unranked. Such data, before they are ranked are called **raw data**.

Definition 2.1.1

Data recorded in the sequence in which they are collected and before they are processed or ranked are called **raw data**.

Consider the following two examples to discuss the concept of raw data

Example 1

Suppose we collect information on the scores of 20 students from Preparatory Year (PY) KSU. The data values, in the order they are collected, are recorded in Table 2.1.

Table 2.1: Scores of 20 students

20	25	18	17	15	19	21	22	27	30
13	15	17	20	25	12	18	13	29	21

The table 2.1 represents quantitative raw data.

Example 2

According to example 1, suppose we ask the same students about their grades as A, B, C and D, after they completed the period study in (PY), and record the values in table 2.2

Table 2.2: Grades of 20 students

A	B	C	D	A	A	B	C	D	A
B	C	D	D	A	B	C	D	A	D

The data in the Table 2.2 is an example of qualitative data.

2.2 Organizing and Graphing Qualitative Data

In this section we will study some methods that used to organize qualitative data set.

A Frequency Table:

The first method to organize a qualitative data is called frequency table, to understand such method, we will review the following example:

Example 1

Assume that a sample of 50 students from the preparatory year (PY) at K.S.U was selected, and those students were asked how they feel about the degree of their satisfaction of the program. The responses of those students are recorded below where (v) means very high satisfaction, (s) means somewhat satisfaction and (n) means no satisfaction.

n	n	n	v	s	n	n	n	v	v
v	s	v	n	n	v	s	n	v	v
v	v	s	n	v	n	s	v	n	n
v	v	s	s	v	v	v	n	s	s
n	s	v	v	v	n	n	s	n	s

We note that twenty of them were very high satisfaction; twelve of them were somewhat satisfied, and eighteen of them were not satisfaction. Table 2.3 (is called a frequency table) listed the type of satisfaction and the number of students corresponding to each category. According to this table, clearly the variable is the type of satisfaction, which is qualitative variable. Note that, each of the students belongs to one and only one of the categories. The number of students who belong to a certain category is called the frequency of that category. A frequency table shows how the frequencies are distributed over various categories.

Refer to example 1 we can view the frequency table 2.3 (for such data) as follow:

Table 2.3:

Type of Satisfaction Variable	Number of Students Frequency
Very high Satisfaction (v)	20
Somewhat satisfaction (s)	12
No satisfaction (n)	18
	Sum=50

Definition 2.2.1

A frequency table for qualitative data lists all categories , names or labels and the number of elements that belong to each of the categories, names or labels.

To construct a frequency table, we will follow the following steps

Construct a Frequency Distribution Table

1. Identify the variable and their categories.
2. Record the categories, names or labels in the first column (or rows) of a table
3. Mark a tally, denoted by the symbol / in the second column, next to the corresponding category, name as label.
4. Record the total of the tallies for each category in the third column, which is called the column of frequencies and usually denoted by f.

Note that, the sum of the entries in frequency column should be equal the sample size.
Constructing a frequency distribution table is also consider in the following example

Example 2

A sample of 50 students from the preparatory year (PY) in K.S.U was selected, and these students were asked how do they feel about the degree of their satisfaction of the program. The responses of those students are recorded below where (v) means very high satisfaction, (s) means somewhat satisfaction and (n) means no satisfaction.

s	s	n	v	s	n	n	n	v	v
v	s	v	n	n	v	s	n	v	v
v	v	s	n	v	s	s	v	n	n
v	v	s	s	v	v	v	n	s	s
s	s	v	v	v	n	n	s	s	s

Construct a frequency distribution table for these data.

Solution: The frequency is constructed in table 2.4

Table 2.4:

Type of Satisfaction	Tally	Frequency (f)
v	//// //// //// ////	20
s	//// //// //// //	17
n	//// //// ///	13
		Sum=50

Note that the variable in this example is the degree of satisfaction of the program in PY. This variable is classified into 3 categories: Very high satisfaction, somewhat satisfaction, and not satisfaction. We record these categories in the first column of table 2.4, then mark a tally, denoted by the symbol / in the second column, next to the corresponding category. Finally, we record the total of the tallies for each category in the third column, which is called the column of frequencies and usually denoted by (f).

Note that we usually delete the column of tally and used only when it is necessary.

B Relative Frequency and Percentage Distributions

The relative frequency is a tool that shows what proportion of the total frequency belongs to the corresponding category; therefore the relative frequency of a category can be calculated by dividing the frequency of that category by the sum of all frequencies. In the column of the relative frequency lists the relatives of all categories.

$$\text{Relative frequency of a category} = \frac{\text{Frequency of that category}}{\text{Sum of all frequencies}}$$

Moreover, multiplying the relative frequency of category by 100% is percentage of a category. In the column of a frequency percentage lists the percentages of all categories.

$$\text{The percentage of a category} = \left(\text{Category Relative Frequency}\right) \cdot 100\%$$

Let us consider an example

Example 3

Determine the relative frequency and percentage table for the data in table 2.4.

Solution: Applying the definition of relative frequency and the percentage of each category we get the following table:

Table 2.5:

Type of Satisfaction	Relative Frequency (f)	Percentage
v	20/50=0.40	(0.40) (100%)=40%
s	17/50=0.34	(0.34) (100%)=34%
n	13/50=0.26	(0.26) (100%)=26%
	Sum= 1.00	**Sum**= 100%

C Graphical Presentation of Qualitative Data

There are many types of graphs that are used to display qualitative data; in this part we will study and graph two of such graphs which they are commonly used to display the qualitative data, these graphs are the **Bar chart** and the **Pie chart**.

1. Bar Chart:

To construct a bar graph (also called a bar chart), we use the following steps

Construct a Bar Graph (Chart)

1. Represent the categories on the horizontal axis (All categories are represented by intervals of the same width).
2. Mark the frequencies on the vertical axis.
3. Draw one bar for each category such that the bar graphs for relative frequency and percentage can be drawn simply by marking the relative frequencies or percentages, instead of the frequencies, on the vertical axis.

Definition 2.2.2

A graph made of bars whose heights represent the frequencies of respective categories is called a bar graph.

Let us consider an example

Example 4

Refer to the table 2.3, we construct bar graph for its data as follows:

Step1: Represent the categories on the horizontal axis and the frequencies on the vertical axis.

Step2: Draw a bar for each category with height equal to its frequency as it is shown by the following figure.

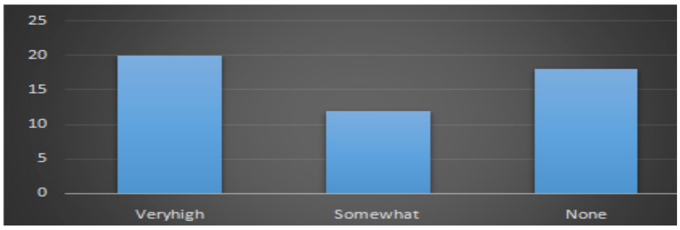


Figure 2.1

2. Pie Chart

The pie chart is one of the most commonly used charts when we need to display percentages, or display frequencies and relative frequencies. In this graph the whole circle (or pie) represents the total sample or population. A pie chart can be drawn by divide the pie into portions that represent the categories.

Definition 2.2.3

A circle divided into portions that represents the relative frequencies or percentages of a population or a sample of different categories is called a pie chart.

To construct a pie graph, we will follow the following steps.

Construct a Pie Graph (Chart)

1. Draw a circle.
2. Find the central angle for each category by the following equation:

   Measure of the central angle = (Relative frequency) $\times 360^{\circ}$
3. Draw sectors corresponding to the angles that obtained in step 2.

Let us consider an example

Example 5

Construct a pie chart for table 2.5

Solution:

Step1: Draw a circle.

Step2: Find the central angle for each category by the equation:

Measure of the central angle= (Relative frequency) $\times 360^{\circ}$

Applying step 2 for each category, we get

For category (v) the measure angle is $(0.40)(360^{\circ})=144^{\circ}$

For category (s) the measure angle is $(0.34)(360^{\circ})=122.4^{\circ}$

For category (n) the measure angle is $(0.26)(360^{\circ})=93.6^{\circ}$

Step2: Draw the sectors corresponding to above angles, then the pie chart for such data is constructed in figure 2.2:

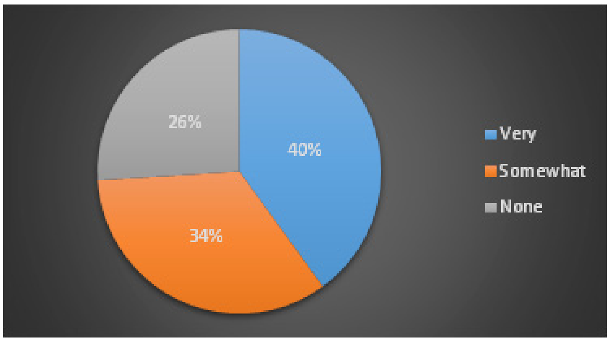


Figure 2.2

Exercises 2.2

1 The table below represent a qualitative data of a sample survey, where A, B, C and D are categories.

A	D	B	C	A	C	B	D	C	B
C	A	A	B	D	D	C	D	A	A
B	C	D	D	B	C	C	A	C	C
B	B	C	B	A	B	C	C	D	B

Then:

a Construct a frequency table for the given data.

b Calculate the relative frequencies and percentages for each category.

c What is the percentage of the elements belongs to category A.

d Draw a bar graph for the frequency table.

e Draw a pie chart for the frequency table.

2 Let be the following percentage table.

Favorite Sport	Percentage of Response
Skating	17 %
Basketball	12 %
Track	23 %
Swimming	13 %
Wrestling	35 %

Then:

a If the total numbers of data is 100 then calculate the relative frequencies for each category.

b Construct a frequency table contains the relative frequencies and percentage.

c Draw the bar graph for the given percentage table.

d Draw the bar graph for the relative frequencies (what do you notice ?).

e Draw a pie chart for the relative frequencies.

f Draw a pie chart for the percentage table (what do you notice ?).

3 The table shows the country represented by the winner of the 10,000 meter run in the Summer Olympic Games in various years.

Year Country	Year Country	Year Country
1912 Finland	1948 Czechoslovakia	1972 Finland
1920 Finland	1952 Czechoslovakia	1976 Finland
1924 Finland	1956 USSR	1980 Ethiopia
1928 Finland	1960 USSR	1984 Italy
1932 Poland	1964 United States	1988 Morocco
1936 Finland	1968 Kenya	1992 Morocco

Which one of the following frequency tables is the correct answer ?

A)

Country	Frequency
Finland	6
Poland	1
Czechoslovakia	2
USSR	2
United States	1
Kenya	1
Ethiopia	1
Italy	1
Morocco	2

B)

Country	Frequency
Finland	7
Poland	1
Czechoslovakia	2
USSR	2
United States	1
Kenya	1
Ethiopia	1
Italy	1
Morocco	2

C)

Country	Frequency
Finland	7
Poland	1
Czechoslovakia	2
USSR	2
United States	1
Kenya	1
France	1
Ethiopia	1
Italy	1
Morocco	2

D)

Country	Frequency
Finland	7
Poland	1
Czechoslovakia	2
USSR	2
United States	1
Ethiopia	1
Italy	1
Morocco	2

4 . The blood types for 40 people who agreed to participate in a medical study were as follows.

O, A, A, O, O, AB, O, B, A, O, A, O, A, B, O, O, O, AB, A, A, A, B, O, A, A, O, O, B, O, O, O, A, O, O, A, B, O, O, A, AB.

Which one of the following frequency tables is the correct answer ?

A) Blood type	Frequency
O	19
A	11
B	5
AB	2

B) Blood type	Frequency
O	18
A	14
B	5
AB	3

C) Blood type	Frequency
O	19
A	13
B	5
AB	3

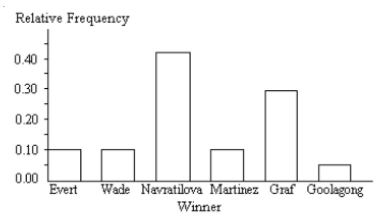
D) Blood type	Frequency
O	20
A	13
B	4
AB	3

5 The table lists the winners of the State Tennis Tournament women's singles title for the years 1986-2005.

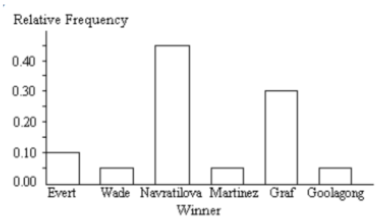
Winner	Frequency	Relative frequency
C. Evert	2	0.10
V. Wade	1	0.05
M. Navratilova	9	0.45
C. Martinez	1	0.05
S. Graf	6	0.30
E. Goolagong	1	0.05

Which one of the following bar graph is the correct answer ?

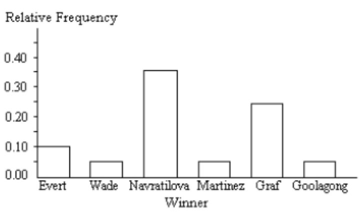
a



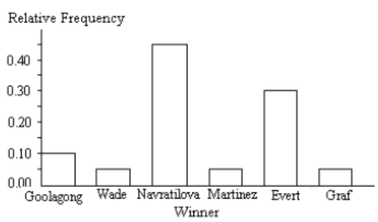
b



c



d

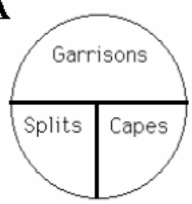


6 The following data give the distribution of the types of houses in a town containing 13,000 houses.

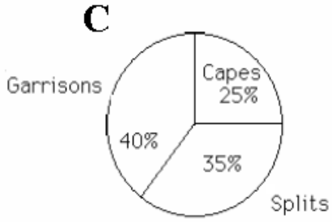
House Type	Frequency	Relative Frequency
Cape	3250	0.25
Garrison	5200	0.35
Split	4550	0.40

Which one of the following pie chart is the correct answer ?

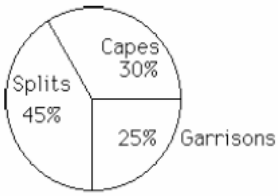
A



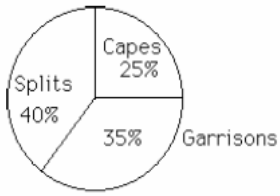
C



B



D



SPSS Statistical Applications

Data entry and graphical representation

In this section, student will learn how to enter raw data into SPSS data file, then how to construct a frequency table, after that how to represent raw data using graphical techniques mentioned in chapter 1.Let's move to practice in example (I) in order to show how SPSS made statistics easy

Example: Suppose the table of data in example1 (responses of 50 KSU-students)

s	s	n	v	s	n	n	n	v	v
v	s	v	n	n	v	s	n	v	v
v	v	s	n	v	s	s	v	n	n
v	v	s	s	v	v	v	n	s	s
s	s	v	v	v	n	n	s	s	s

Where (v) means very satisfied, (s) means somewhat and (n) means not satisfied.

Need to:

1 Enter data using SPSS data file, let the variable name be "response" and save the filewith the name"data2.1"

2 construct a frequency table for the given data

3 draw a bar graph for data collected within response variable by showing percentages.

4 draw a pie graph for data collected within response variable by showing percentages.

Solution:

1. **To enter raw data, open a new SPSS data file by double clicking the SPSS program icon, then:**

a Go to variable view as shown in chapter 1, in the name field write response.

b Type of variable will be string and measurement level will be ordinal

c Go to data view and start to enter letters (v, s or n) under the variable "response"with a systematic order (i.e. start with the first row, and the next and so on).

d From file menu, select save and choose the destination, in name box write data2.1, then ok.You will get the file as in figure 1

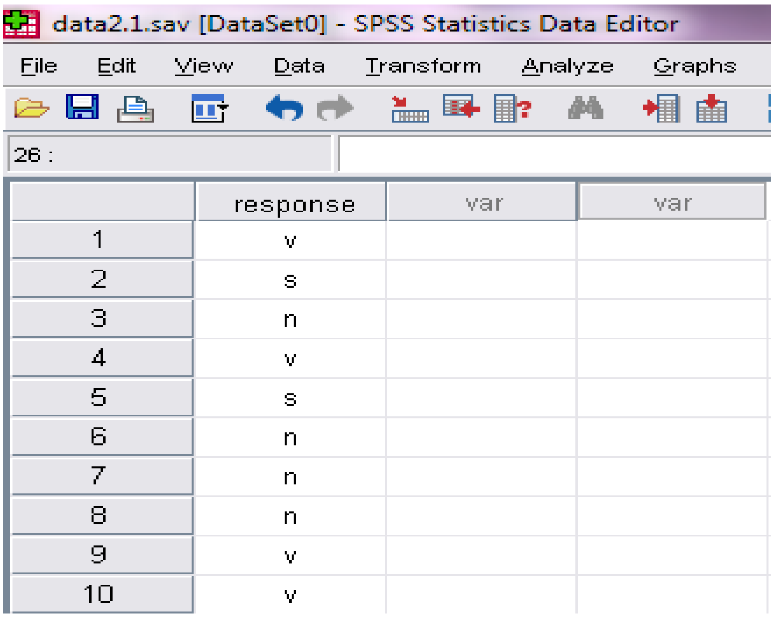


Figure1. Snapshot of the data file created

2. To construct a frequency table, follow the command:

Analyze →Descriptive statistics →Frequencies →select variable "response" →press on the arrow to move it into variables box → OK.(see figure 2)

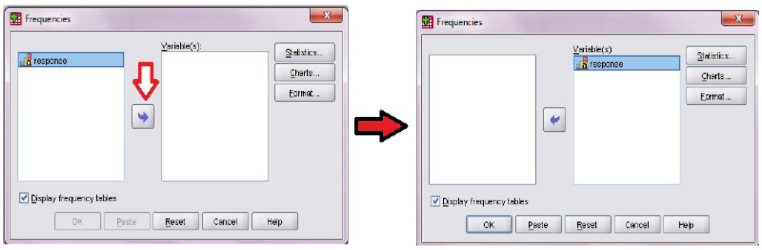


Figure2. Frequency dialog box for response variable

Results will appear in an output file, and we get the following frequency table:

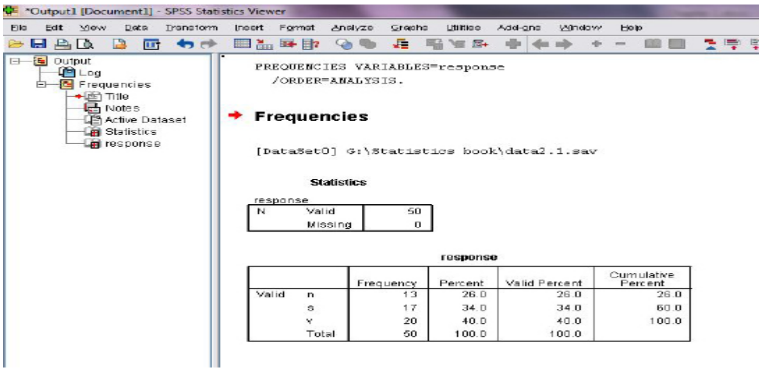


Figure3. Output window with frequencies results

The output displays that:

Response	Frequency	Percentage
n	13	26.0%
s	17	34.0%
v	20	40.0%
Total	50	100.0%

Compare the resulted frequency table with the one in section 2.1, it gives the same results.

3. In order to draw a bar graph, follow the command:

> Graphs→Legacy dialogs→Bar→Simple→define→select variable "response" as category axis→select % of cases to represent bars → OK.(see figures4 and 5)

After drawing the bar graph, one can show percentages on bars, by double clicking the resulted graph on output window to get chart editor and press on blue icon as shown in figure 6.

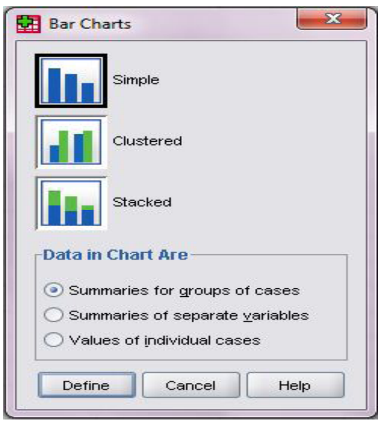


Figure4. Bar chart types dialog

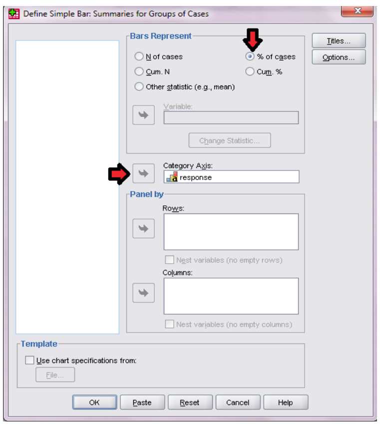


Figure5. Define category axis and representation

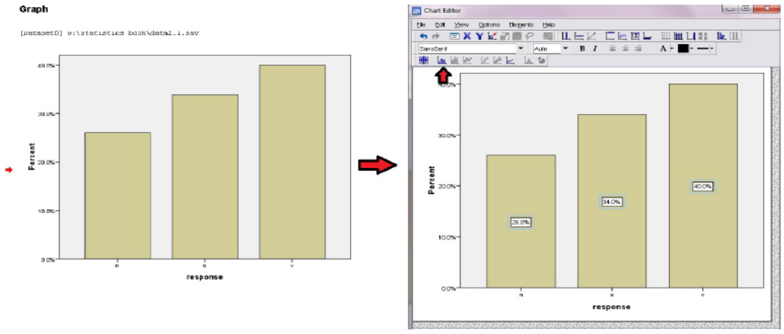


Figure6. Chart editor to show percentages on bars

4. In order to draw a pie graph, follow the command:

Graphs →Legacy dialogs →Pie →choose Summarizes for group of cases →define → Define slice by “response → select % of cases to represent bars →OK .(see figures7 and 8)

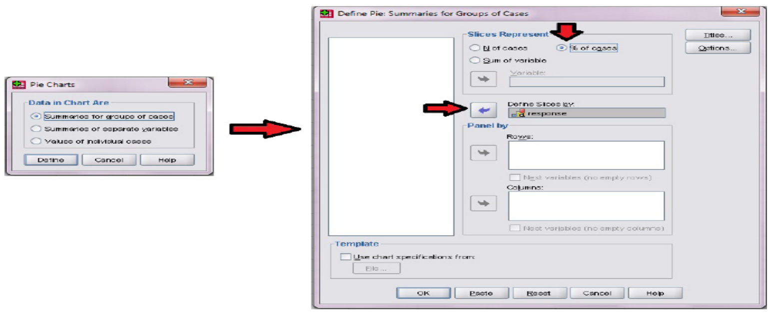


Figure7. Pie charts definition dialog box

The resulted pie graph after editing and showing percentages as the same done in bar graph will be as in figure 8.

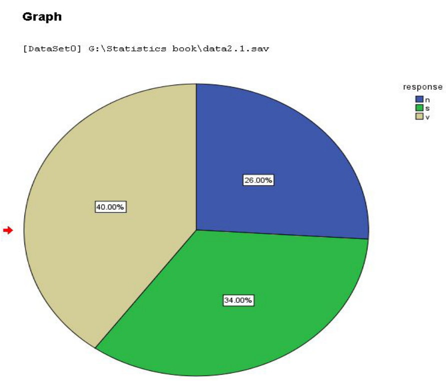


Figure8. Pie graph for response variable

2.3 Organizing and Graphing Quantitative Data

In this section we want to learn how to group and display the second type of data which is the quantitative data.

A Frequency Distributions

Before discussing how construct a frequency distribution table, it is necessary to explain some related concepts, all of these concepts will be discussed by considering an example. Assume that we have the following data table that shown the number of cell-phones sold by email.

Cell-phone Sold by email

SII Sold	Frequency
2-5	7
6-9	5
10-13	2
14-17	6
18-21	3

Variable

Fourth Class

Lower Limit

Upper Limit

According to above table, you can easily see that the first column lists represent the cell-phone sold as interval (and we call class), this column lists are called **classes** and each class consist of two numbers, the first number is called **lower limit** and the second number is called the **upper limit**. The next column represent the number of sold phones such that each number belongs to one and only one interval, these listed numbers are called **frequency** denoted by (f) where each frequency gives the number of values that belong to each class.

Remark

A class is an nonoverlaping interval that includes all values that fall within two numbers.

Definition 2.3.1

Data presented in the form of a frequency distribution table are called grouped data.

Remark

The classes always represent a variable and they are nonoverlapping; that is, each value in the variable belongs to one and only one class.

For our example we have five non overlapping classes. Each class has a lower limit and an upper limit. The value 2 in the first class (2-5) is the lower limit and the value 5 is the upper limit. Therefore, the values 2, 6, 10, 14, and 18 give the lower limits, while the values 5, 9, 13, 17, and 21, are the upper limits of the five classes, respectively.

The above example illustration a frequency distribution table for quantitative data. For ungrouped quantitative data there is another type of table called frequency table and it is used when data list individual values. For example let us consider the following data:
8, 9, 8, 6, 8, 6, 6, 6, 9, 9, 9, 6, 8, 7, 7, 8, 9, 6, 8, 6, 9, 6, 9, 6, 8, 7.
We can then download this data in the following frequency table:

frequency table for quantitative data

Data	Frequency
6	9
7	3
8	7
9	7
Sum	**26**

Class Boundaries

Class Boundaries are similar to class limit, where each class boundaries divided into lower boundary and upper boundary. However, the upper boundary limit for any class always the same of the lower boundary limit of previous class.
The following steps explain how to find class boundaries.

Construct a Frequency Distribution

1. Find the difference between the lower limit of any class and the upper limit of the previous class. Denoted by (d), where
d=lower limit of class (n) - upper limit of class (n-1)

2. Divide the difference (d) by 2

3. Subtract ($d/2$) from each lower limit and add (d/2) for each upper limit.

When the limits of classes are whole numbers that is {0,1,2,3...} readily subtract and add the value (**0.5**) to the lower and upper limits respectively.

A frequency distribution for quantitative data lists all classes and the number of values belonging to each class.

Definition 2.3.2

Data presented in the form of a frequency distribution table are called grouped data.

The rule of class boundaries and completely different from the rule of class limit. This difference will appear when we represent the data diagrammatically. Generally class boundaries limits used to find the class width as follows:

Class width = Upper boundary limit – Lower boundary limit

Let us consider an examples

Example 1

Find the class boundary of the class 40 – 44

Solution:

Lower boundary limit= $40 - 0.5 = 39.5$

Upper boundary limit= $44 + 0.5 = 44.5$

The class boundary is $39.5 - 44.5$

Example 2

Find the class width of the class 40 – 44

Solution: The class boundary is $39.5 - 44.5$. Therefore

The class width is $44.5 - 39.5 = 5$

Midpoints

Dividing the sum of the two boundaries (or the two limits) of a class by 2 gives the class midpoint. So, we can calculate the class midpoint as:

$$x_m = \frac{\text{Lower boundary} + \text{Upper boundary}}{2}$$
$$= \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

Example 3

Find the class midpoint of the class 40 – 44.

Solution: The lower limit is 40 and the upper limit is 44, then the midpoint of this class is:

$$x_m = \frac{40 + 44}{2} = 42 \text{ or } x_m = \frac{39.5 + 44.5}{2} = 42.$$

Constructing a Frequency Distribution

When constructing a frequency distribution table from a data set, we need to apply the following steps.

Construct a Frequency Distribution Table

1. If the number of classes is not given, decide on a number of classes to use (take a natural number as default number).
2. Find the class width: Determine the range (largest value-smallest value) of the data and divide this by the number of classes. Round up to the next convenient number (if it's a whole number, also round up to the next whole number).
   Find the class limits: Take a value less than the minimum data entry as the lower limit of the first class (you can use the minimum data entry as the lower limit of the first class). To get the lower limit of the next class, add the class width to the lower limit of the previous class. Continue until you reach the last class. Then find the upper limits of each class (since the classes cannot overlap, and occasionally your data will include decimal numbers, remember that it's fine for the upper limits to be decimals).
3. Count the number of data entries for each class, and record the number in the row of the table for that class. (The book recommends using "tally" marks to count).
4. The ascending cumulative frequency for a class in the distribution table is equal to the sum of all frequencies until the upper bound of this class in the same distribution table. We note that the ascending cumulative frequency for last class equals to the sum of all frequencies.

Let us consider an example

Example 4

The following data table represent the scores of 50 students in statistics course. Use the this data to

a. Construct the frequency distribution table by using 12 classes

b. Construct the relative frequency distribution table

c. Find the percentages

49	57	38	73	81	74	59	76	65	69
54	56	69	68	78	65	85	49	69	61
48	81	68	37	43	78	82	43	64	67
52	56	81	77	79	85	40	85	59	80
60	71	57	61	69	61	83	90	87	74

Solution:

Step1: The largest score is 90, the lowest one is 37, so

$$\text{The range} = 90\text{–}37 = 53$$

Step2: Divide the range by the number of classes, which is twelve

$$\frac{53}{12} = 4.4$$

Step3: Round up the number in step 2 to the next whole number that is 5. Thus, the class width is 5.

Step4: Identify the first class in the distribution table that contains the lowest value (we can use the lowest value of the data as a lower bound for the first class). Each class has a lower limit and an upper limit. If we choose the first class limit 35–39, then we have the second class limit is 40–44, the third class limit is 45–49. Continue until the last class that contains the largest value 90, so the last class limit is 90–94.

Step5: Determine the class boundaries according to the scoop was explained. So we have for our example: the first class boundary is 34.5 – 39.5, then we have the second class boundaries is 39.5–44.5. Continue until the last class boundaries is 89.5–94.5.

Step6: Use the symbol (/) instead of each value of the raw data opposite the class which the value belongs to it in the next column. The number of tallies in each class is called the frequency of the class that denoted by f recorded in the next column. So, we construct the following frequency distribution table.

Step7: The ascending cumulative frequency for the first class is 2, and for the second class is 5, Continue until the last class we have the ascending cumulative frequency for it is 50.

Class limit	Class boundaries	Midpoint	Tally	Frequency f	Relative frequency	Percent-age	Ascending cumulative frequency
35–39	34.5–39.5	37	//	2	0.04	4%	2
40–44	39.5–44.5	42	///	3	0.06	6%	5
45–49	44.5–49.5	47	///	3	0.06	6%	8
50–54	49.5–54.5	52	//	2	0.04	4%	10
55–59	54.5–59.5	57	~~////~~ /	6	0.12	12%	16
60–64	59.5–64.5	62	~~////~~	5	0.10	10%	21
65–69	64.5–69.5	67	~~////~~ ////	9	0.18	18%	30
70–74	69.5–74.5	72	////	4	0.08	8%	34
75–79	74.5–79.5	77	~~////~~	5	0.10	10%	39
80–84	79.5–84.5	82	~~////~~ /	6	0.12	12%	45
85–89	84.5–89.5	87	////	4	0.08	8%	49
90–94	89.5–94.5	92	/	1	0.02	2%	50
Total				**50**	**1**	**100 %**	

Note that, when you construct a distribution frequency table usually suffice to provide the following columns:

- No.of Class.
- Class Boundary.
- Midpoint.
- Frequency.
- Ascending Cumulative Frequency

Thus, we can offer the above table as follows:

No. of Class	Class boundaries	Midpoint	Frequency f	Ascending cumulative frequency
1	34.5–39.5	37	2	2
2	39.5–44.5	42	3	5
3	44.5–49.5	47	3	8
4	4.5–54.5	52	2	10
5	54.5–59.5	57	6	16
6	59.5–64.5	62	5	21
7	64.5–69.5	67	9	30
8	69.5–74.5	72	4	34
9	74.5–79.5	77	5	39
10	79.5–84.5	82	6	45
11	84.5–89.5	87	4	49
12	89.5–94.5	92	1	50
Total		**50**		

B Graphing Grouped Data

Grouped data can be displayed in a histogram, a polygon or ogive. In this part we will learn how to construct such graphs.

1. Histogram

Histogram is a graphical technique that used for representation of a frequency, a relative frequency, or a percentage of frequency distribution table of grouped data.

Definition 2.3.3

A histogram of grouped data in a frequency distribution table with equal class widths is a graph in which class boundaries are marked on the horizontal axis and the frequencies, relative frequencies, or percentages are marked on a vertical axis.

Note that in the histogram, the bars are drawn adjacent to each other.

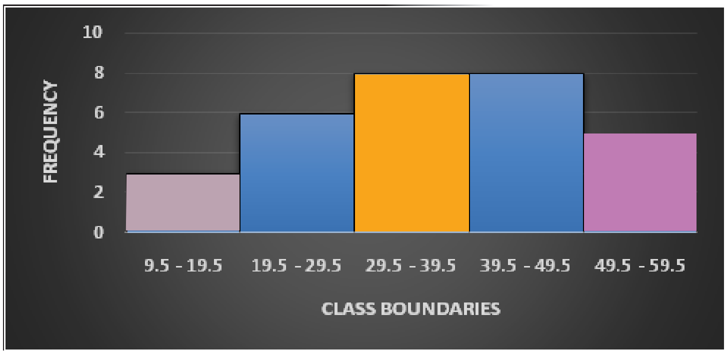
The following example illustrates how we can graph a histogram

Example 5

Use the table below to represent the data by a histogram and the relative frequency histogram.

Class	Class boundary	Frequency (f)
10–19	9.5-19.5	3
20–29	19.5 – 29.5	6
30–39	29.5 – 39.5	8
40–49	39.5 – 49.5	8
50–59	49.5 – 59.5	5

Solution: We find the boundaries for all classes and use these boundaries with its frequencies



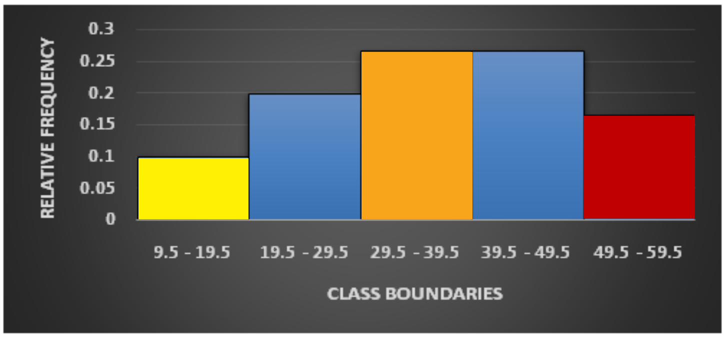
Using the formula of relative frequency that illustrated in section 2.1

$$\text{Relative frequency} = \frac{\text{Frequency of the class}}{\text{Sum of all frequencies}} = \frac{f}{\sum f}$$

So, the relative frequency distribution table is as follows:

Class boundaries	Relative frequencies
9.5 – 19.5	$3/30 = 0.10$
19.5 – 29.5	$6/30 = 0.20$
29.5 – 39.5	$8/30 = 0.267$
39.5 – 49.5	$8/30 = 0.267$
49.5 – 59.5	$5/30 = 0.166$
	Sum = 1.00

The relative frequency histogram is:



2. Polygons

A polygon is another type of graphs that can be used to represent grouped quantitative data. To draw a frequency polygon we plot the points (class midpoint, frequency) and connect these points by line segments.

Definition 2.3.4

A frequency polygon is a graph that displays the data by using line segments that connect points plotted for the frequencies at the midpoints of the classes.

Note that, we add one midpoint with zero frequency before the first class and after the last class midpoint to closed the graph.

To construct a frequency polygon from a data set, we need to apply the steps below.

Construct a Polygon

1. Find the class midpoints.
2. Mark the class midpoints on the horizontal axis.
3. Mark the frequencies on the vertical axis.
4. Plot the points (Class midpoint, Frequency).
5. Connect these points by straight lines segments.
6. Close the polygon from the left by taking a class before the first class and take its frequency equal to zero, and close the polygon from the right by taking a next class with frequency also zero.

Let us consider an example

Example 6

Draw the frequency polygon for the given distribution table in example 5

Solution:

Step1: Find the class midpoints as follows:

Class	Midpoint
10 – 19	14.5
20 – 29	24.5
30 – 39	34.5
40 – 49	44.5
50 – 59	54.5

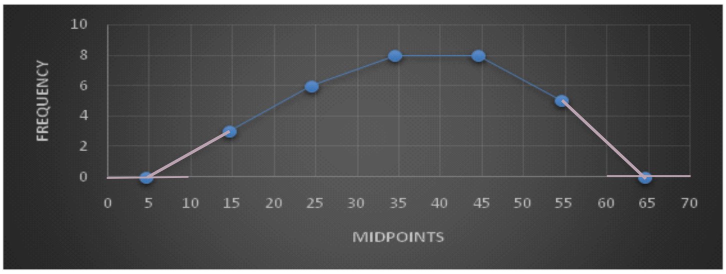
Step2: Mark the class midpoints on the horizontal axis.

Step3: Mark the frequencies on the vertical axis.

Step4: Plot the points (Class midpoint, Frequency).

Step5: Connect these points by straight lines segments.

Step6: Close the polygon from the left and right.



3. Ascending Cumulative Frequency Curve (Ogive)

The third type of graphs that can be used to represent the cumulative frequencies for the classes is cumulative frequency graph or ogive.

The cumulative frequency is the sum of the frequencies accumulated up to the upper boundary of a class boundaries in the distribution table.

Definition 2.3.5

An ogive is a curve drawn for the ascending cumulative frequency of grouped data in a distribution table by first joining plotting dots marked above the upper boundaries of classes at heights equal to the ascending cumulative frequencies of respective classes, then joining these points by smooth curve.

When constructing a cumulative frequency curve from a grouped data set, we need to apply the following steps .

Construct Ascending Cumulative Frequency

1. Mark the upper boundaries on the horizontal axis.
2. Mark the ascending cumulative frequencies in the vertical axis.
3. Plot the points of the coordinates (upper boundary, ascending cumulative frequency).
4. Connect each two adjacent points with a curve (Smoothly).
5. Close the curve from the left to the lower limit of first class boundary.

Let us consider an example

Example 7

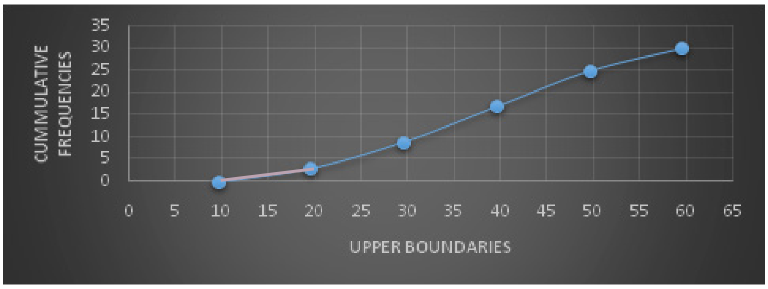
Draw the ogive for the frequency distribution table in example 5

Solution: Step1: Calculate the ascending cumulative frequencies and record they in their own column.

Upper boundaries	Cumulative Frequency
Less than 19.5	3
Less than 29.5	9
Less than 39.5	17
Less than 49.5	25
Less than 59.5	30

Step2: Mark the upper boundaries on the horizontal axis.

Step3: Mark the cumulative frequencies in the vertical axis.
Step4: Plot the points (upper boundary, ascending cumulative frequency).
Step5: Connect each two adjacent points with a curve (Smoothly).
Step6: Close the curve from the left to the lower limit of first class boundary.



4. Stem-and-Leaf Display

A stem and leaf is another type of table for displaying quantitative data. A stem and leaf display of quantitative data divides each value into two portions, a stem and leaf. The leaves for each stem are shown separately in a display.

For constructing stem-and-leaf from a data set, we follow the following steps

Construct a Stem and Leaf

1. Construct a table with two columns
2. The first column contains the first digit, which is called the stem
3. The second column contains the second digit, which is called the leaf
4. Arrange leafs in an increasing order.

Let us consider an example

Example 8

Construct a stem-leaf display for the following data: (Scores for 30 students).

75	69	83	52	72	84	80	81	77	96
61	64	65	76	71	79	86	87	72	79
72	87	68	92	93	50	57	95	92	98

Solution:

Step1: The first part contains the first digit (the tens), which represents the stem.

Step2: The second part contains the second digit, which represents the leaf.

We find for the first student, score 7 represent the stem and 5 is the leaf. Thus we have the following table of the data given.

Stem	Leaf
5	2 0 7
6	9 1 4 5 8
7	5 2 7 6 1 9 2 9 2
8	3 4 0 1 6 7 7
9	6 2 3 5 2 8

Finally, arrange the leafs in an increasing order, to get

Stem	Leaf
5	0 2 7
6	1 4 5 8 9
7	1 2 2 2 5 6 7 9 9
8	0 1 3 4 6 7 7
9	2 2 3 5 6 8

Example 9

Let be the following data
45,50,55,60,65,70,75,47,51,56,61,66,71,76,48,52,57,62,67,72,77,49,53,58,63,68,73,78,49,54,59,64,68,74,49,51,55,61,68,71,51,56,61,69,71,52,56,62,66,72,53,57,62,67,72,54,58,63,67,74,58,63,68,58,64,68,59,64,69,55,64,69,56,64,68,61,61,62,62,63.

Then:

a) Make this data in frequency distribution table with Class boundaries width equal to 5 and containing columns of the following:

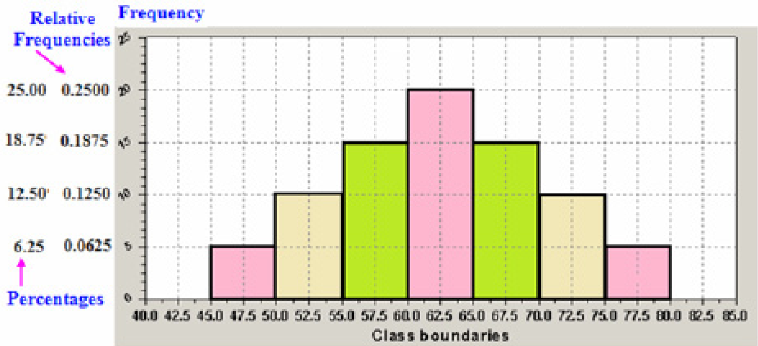
1 Class boundaries
2 Midpoints
3 Frequencies
4 Relative frequencies.
5 Percentages
6 Ascending cumulative frequencies

b) Draw the histogram for the frequencies in the distribution table.
c) Draw the histogram for the relative frequencies in the distribution table.
d) Draw the histogram for the percentage frequencies in the distribution table.
e) Draw the polygon for the frequencies in the distribution table.
f) Draw the polygon for the relative frequencies in the distribution table.
g) Draw the polygon for the percentage in the distribution table.
h) Draw the ogive of data in the distribution table.

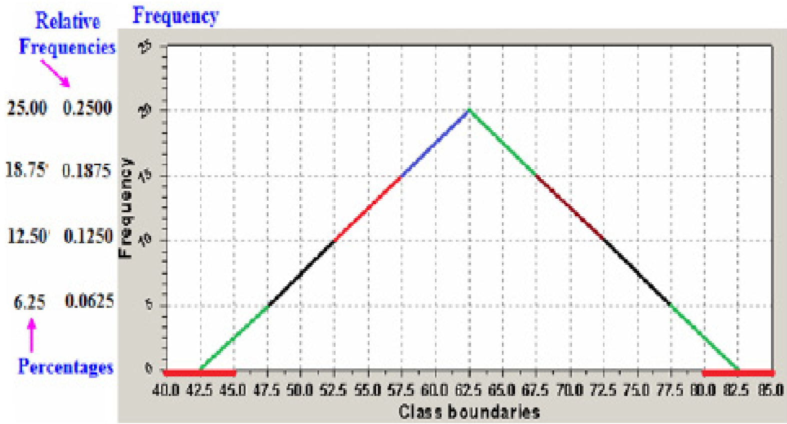
Solution:
The frequency distribution table with Class boundaries width equal to 5 is for the given data as following:

No of Class	Class Boundaries	Midpoint	Frequency	Relative Frequencies	Percentages	Ascending Cumulative Frequency
1	45-50	47.5	5	0.0625	6.25%	5
2	50-55	52.5	10	0.1250	12.50%	15
3	55–60	57.5	15	0.1875	18.75%	30
4	60–65	62.5	20	0.2500	25.00%	50
5	65–70	67.5	15	0.1875	18.75%	65
6	70–75	72.5	10	0.1250	12.50%	75
7	75–80	77.5	5	0.0625	6.25%	80
Sum			**80**	**1**	**100%**	

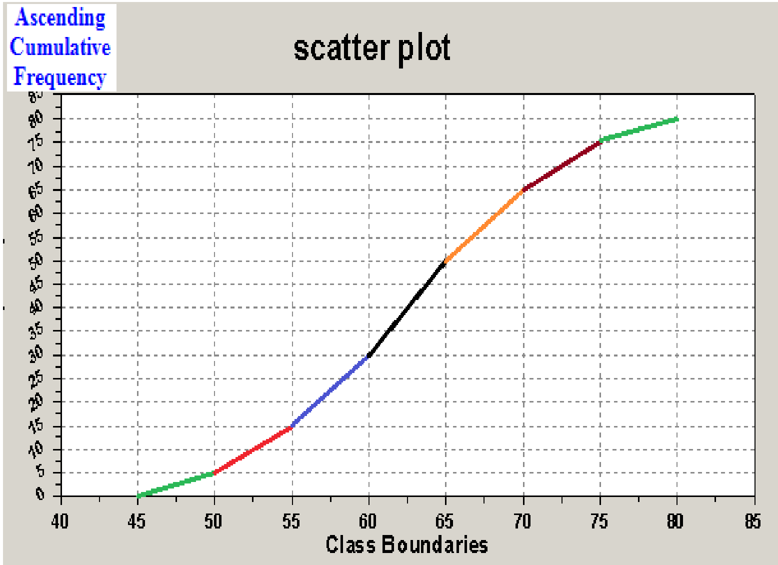
Note that we have the range of given data equal to 77.5- 47.5 = 30
The histogram, relative histogram and percentage histogram for the given frequency distribution table are as in the following figure:



The polygon, relative polygon and percentage polygon for the given frequency distribution table are as in the following figure:



The ogive for the given frequency distribution table is as in the following figure:



Exercises 2.3

1 For the following frequency distribution table

Class	Frequency
0 – 3	23
4 – 7	40
8 – 11	28
12 – 15	6
16 – 19	3

a Find the class boundaries and class midpoints.

b What is the width of the class.

c Construct the relative frequency and percentage distribution table.

d What percentage of data with value equal to 8 or more.

e What percentage of data with value equal to 11 or less.

2 For the following data

45	49	48	41	54	46	44	42	48	53
51	53	51	48	46	43	52	50	54	47
44	45	50	49	50					

a Construct the frequency table.

b Calculate the relative frequencies for all classes.

c Construct a histogram for relative frequency distribution.

d Construct a polygon for the relative frequency distribution.

3 For the following distribution

Class	Frequency
0 – 3	20
4 – 7	40
8 – 11	30
12 – 15	7
16 – 19	3

a Construct a cumulative frequency distribution table.

b Calculate the cumulative relative frequencies and cumulative percentages for all classes.

c Find the percentage of these who possess 11 or less than.

d Draw an ogive for the cumulative percentage distribution.

e Use the ogive, find the percentage who possess 9 or less than.

4 For the following distribution

Class	Frequency
18 – 30	8
31 – 43	22
44 – 56	14
57 – 69	6

a Construct a cumulative frequency distribution table.

b Calculate the cumulative relative frequencies and cumulative percentages for all classes.

c What percentage of data with value equal to 44 or more.

d Draw an ogive for cumulative percentage distribution.

e Using the ogive, find the percentage of 40 or less than.

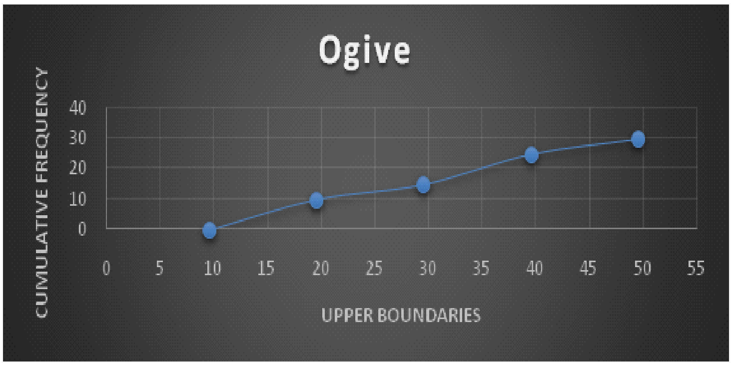
5 Construct a stem-leaf display for the data:

15	08	23	21	04	17	31	22	31	06
05	06	14	17	16	25	27	03	31	08

6 Construct a stem-leaf display for the data:

10	50	65	33	48	5	11	23	39	26
26	32	17	7	15	19	29	43	21	22

7



a Find the total frequency of the above distribution.

b Find the value with cumulative frequency less than or equal 15.

c Find the cumulative frequency of the value 25.

8 A random sample of 30 high school students is selected. Each student is asked how much time he or she spent on the internet during the previous week. The following times (in hours) are recorded.

6	14	8	11	8	6	8	7	5	11
9	7	7	6	9	8	5	5	10	7
5	7	14	9	6	10	6	9	8	7

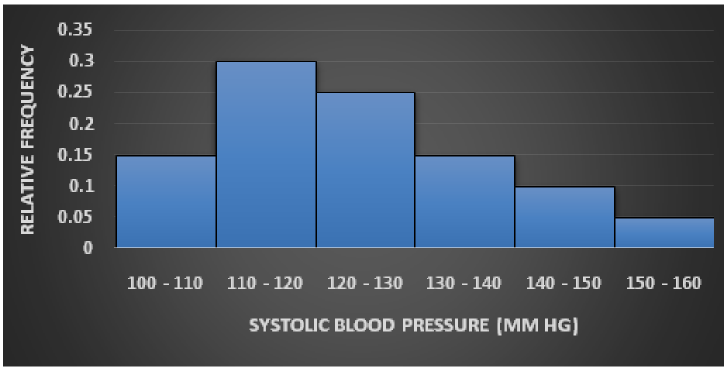
a Construct a frequency histogram for this data.

b Construct a frequency polygon for this data.

9 The June precipitation amounts (in inches) for 40 cities are listed below. Construct a frequency distribution and a relative frequency distribution using eight classes.

2.0	3.2	1.8	2.9	0.9	4.0	3.3	2.9	3.6	0.8
3.1	2.4	2.4	2.3	1.6	1.6	4.0	3.1	3.2	1.8
2.2	2.2	1.7	0.5	3.6	3.4	1.9	2.0	3.0	1.1
3.0	4.0	4.0	2.1	1.9	1.1	0.5	3.2	3.0	2.2

10 A nurse measured the blood pressure of each person who visited her clinic. Following is a relative frequency histogram for the systolic blood pressure readings for those people aged between 25 and 40. Use the histogram to answer the question 11-13. The blood pressure readings were given to the nearest whole number.



a Approximately, what percentage of the people aged 25-40 had a systolic blood pressure reading between 110 and 119 inclusive.

b Approximately, what percentage of the people aged 25-40 had a systolic blood pressure reading between 110 and 139 inclusive.

c Approximately, what percentage of the people aged 25-40 had a systolic blood pressure reading greater than or equal to 130.

11 The table below contains the frequency and relative frequency distributions for the ages of the employee in a particular company department. Construct a relative frequency polygon for these data.

Ages (years)	Frequency	Relative frequency
20 – 30	6	0.375
30 – 40	3	0.1875
40 – 50	4	0.25
50 – 60	2	0.125
60 – 70	1	0.0625

12 Let be the following data
60,33,36,38,47,45,49,46,25,13,39,48,20,46,12,29,38,09,57,51,25,27,29,21,20,46,34,39,39,57,18,44,36,30,18,39,36,27,33,48,56,47,20,50,58,37,27,32,36,29,37,05,45,47,55,55,35,60,26,18.

Then

a Make this data in frequency distribution table with Class boundaries width equal to 5 and containing columns of the following:

1 Class boundaries 2 Midpoints
3 Frequencies 4 Relative frequencies
5 Percentages 6 Ascending cumulative frequencies

b Draw the histogram for the given frequency distribution table

c Draw the relative histogram for the given frequency distribution table

d Draw the percentage histogram for the given frequency distribution table

e Draw the polygon for the given frequency distribution table

f Draw the relative polygon for the given frequency distribution table

g Draw the percentage polygon for the given frequency distribution table

h Draw the ogive for the given frequency distribution table

13 The class boundary of the class 20 – 50 is

A) 19.5-20.5 B) 19.5-49 C) 19.5-50.5 D) 20-50.5.

14 The class width of the class 70 – 79 is

A) 9 B) 10 C) 79 D) 74.5

15 The class midpoint of the class 40 – 44 is

A) 40 B) 44 C) 42 D) 44.5

16 The data in the following table reflect the amount of time 40 students in a section of Statistics 140 spend on homework each day

Homework time (minutes)	Relative frequency
0-14	0.05
15-29	0.10
30-44	0.25
45-59	
60-74	0.15
75-89	0.05

The value of the missing entry is

A) 40% B) 16 C) 0.40 D) 1.

SPSS Statistical Applications

Frequency table for grouped data and drawing histogram

In this section, student will learn how to construct a frequency table for grouped data. inorder to do this, student should know how to organized data in intervals as illustrated in section 2.2. Let's move to practice.

Example: Suppose we have the following data represent ages (in years) of a sample of (30) visitors to a Mall:

10	20	30	41	51	15	21	22	41	52
18	22	33	42	53	22	34	42	55	27
35	45	56	28	36	46	38	47	39	48

Need to:
1) Enter data using 5 intervals, let the variable name be "age_class" and save the file with the name"data2.2"

2) construct a frequency table for the given data with the following five classes:

(10 - 19), (20 - 29),(30 - 39),(40 - 49),(50 - 59).

3) draw a histogram for the grouped data collected within age_class variable.
4) draw a stem and leaf for the collected raw data.

Solution:

1) To enter raw data, open a new SPSS data file by double clicking the SPSS program icon, then:

* Go to variable view, in the name field write data.
* Type of variable will be numeric and measurement level will be scale
* Go to data view and start to enter raw data that represents ages of visitors to a mall under the variable "data"with a systematic order (i.e. start with the first row, and the next and so on).
* From file menu, select save and choose the destination, in name box write data2.2, then ok
* From variable view, go to decimals field and reduce it to zero instead of 2
* In data view, right click on variable to get a small menu, choose sort ascending to get data in ordered form.You will get the file as in figure 1

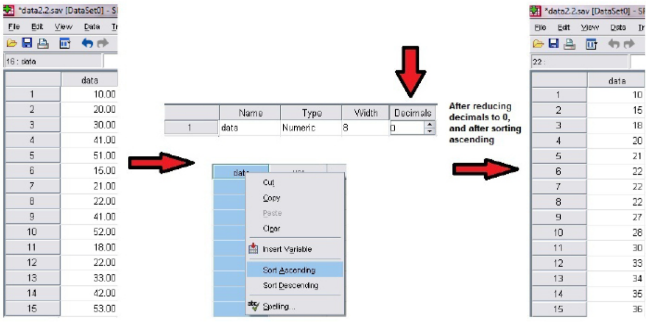


Figure1. Snapshot of data view of data2.2 file after sorting and reduction of decimals

2) To construct a frequency table, need first to define intervals using "Transform" menu follow the command:

Transform →(step1): Recode into different variables→(step2): choose "data" variable →(step3): in output variable box write name "age_class" →press on change → press on Old and new values to make classes→ (step4) enter the range of old data and new values(e.g. from 1 to 5) →press add each time → continue. (see figure 2).

Figure 2. displays the steps of transforming the old values to the new values in order to get the mentioned classes:

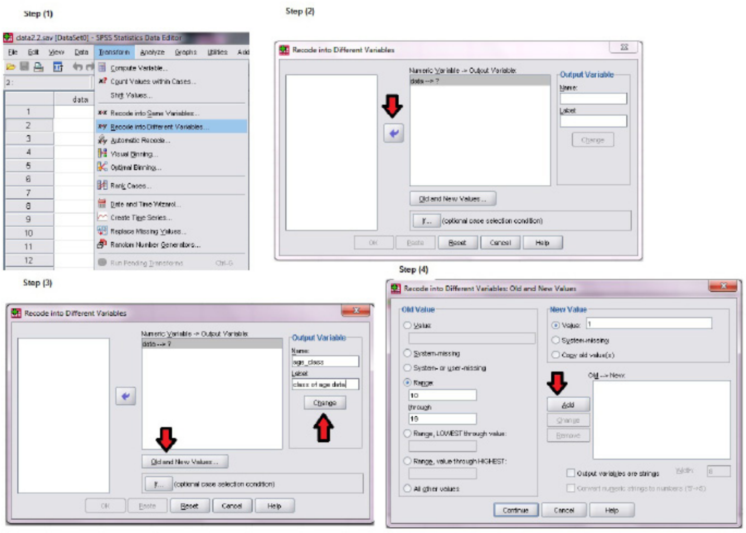


Figure2. Steps of recoding values of "data" variable into "age_class" variable

Adding labels to the new values, in order to plot the histogram is needed, this can be done using the "value label" field in variable view of SPSS data file as to give number 1 label as (10-19) and so. See figure 3

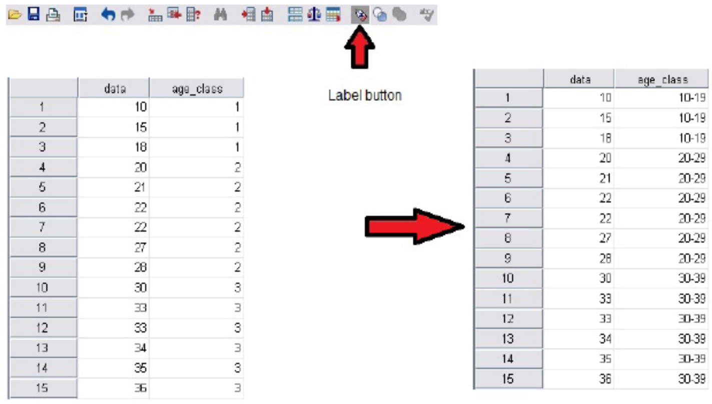


Figure4. label button to view new data classes for the variable "age_class"

Now, to construct the frequency table for grouped data or classes follow the same command as we did in statistical application of section 2.1

Analyze →Descriptive statistics →Frequencies →select variable "age_class" →press on the arrow to move it into variables box → OK.(see figure 5)

We get,

class of age data

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10-19	3	10.0	10.0	10.0
	20-29	6	20.0	20.0	30.0
	30-39	8	26.7	26.7	56.7
	40-49	8	26.7	26.7	83.3
	50-59	5	16.7	16.7	100.0
	Total	30	100.0	100.0	

Figure5. frequency table for age_class variable

3) **To draw a histogram for the variable "age_class", follow the command:**

Graphs →Legacy dialogs →Histogram →choose "age_class" in variable field →OK.(see figures6)

We get,

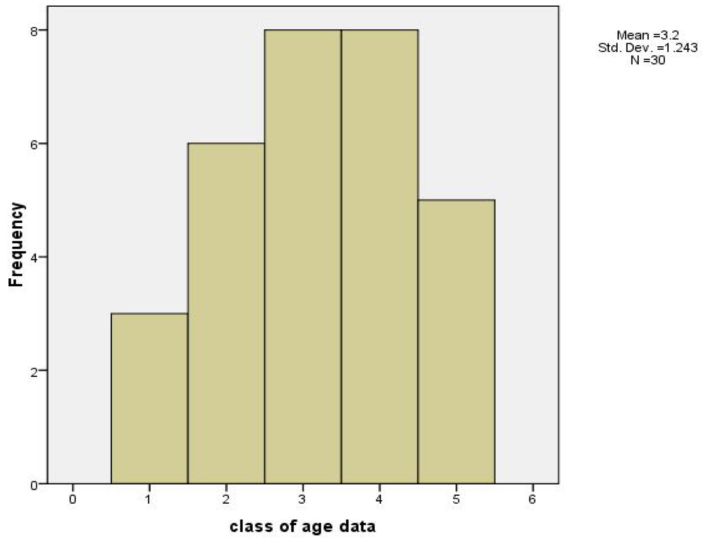


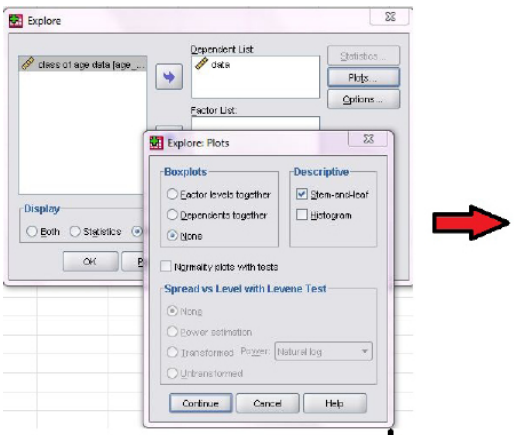
Figure 6. Histogram of age_class variable

Note that, the number 1- 5 represents the classes (1 → 10-19)... (5 → 50-59). One can get the polygon by drawing a line crossing the midpoint of each bar. Then close the polygon from the first by taking a previous class its frequency is zero, and close the polygon from the last by taking a next class which its frequency also zero.

4) To draw a stem and leaf for the variable "age_class", follow the command:

Analyze →Descriptive statistics →Explore→choose "data" in dependent list field Plots →choose stem and leaf → continue → OK.(see figures7)

The resulted plot will be as follows:




```
data Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     1.00        1 .  0
     2.00        1 .  58
     4.00        2 .  0122
     2.00        2 .  78
     4.00        3 .  0334
     4.00        3 .  5689
     4.00        4 .  1122
     4.00        4 .  5678
     3.00        5 .  123
     2.00        5 .  56

 Stem width:        10
 Each leaf:       1 case(s)
```

Figure 7.Stem and Leaf plot for data variable

CHAPTER 3

Numerical Descriptive Measures

OBJECTIVES

1. Describe data using measures of central tendency, such as the mean, median, mode, and midrange.
2. Summarize data using measures of variation, such as the range, variance, and standard deviation.
3. Determine the position of a data value in a data set using various measures of position, such as percentiles, deciles and quartiles.
4. Use the techniques of exploratory data analysis including box plots and five-point summary statistics to discover various aspects of data.

After we discussed in chapter two how to group data using frequency distributions, and how to represent the data set by using graphs, it is necessary to use numerical measures to display a data set. In this chapter, we want to study the numerical measures such as ones of central tendency, variation, measures of position and the Box-whisker.

3.1 Measures of Central Tendency

Remark

The values of central tendency does not change, whether the data belong to the population or to the sample. Therefore, when calculating the central tendency is not important to look at the data source (data of population or data of sample).

A measure of central tendency is very important tool that refer to the centre of a histogram or a frequency distribution curve. In This section we will discuss three measures of central tendency and learn how to calculate it. Such measures are the **mean**, the **median**, and the **mode** for the two cases (grouped and ungrouped data sets).

A The Mean

The most commonly used measure of central tendency is called mean (or the average). Here the main of interest is to learn how to calculate the mean when the data set is in type of ungrouped (raw data).

1. The Mean for Ungrouped Data:

In simple words the mean can be calculated for ungrouped data set as it is illustrated by the following definition

Definition 3.1.1

The mean for an ungrouped data is obtained by dividing the sum of all values by the number of values in that data set. Thus,

Mean for population data: $\mu = \frac{\sum x}{N}$

Mean for sample data: $\bar{x} = \frac{\sum x}{n}$

Where $\sum x$ is the sum of all values, N is the population size, and n is the sample size, μ is the population mean, and $\bar{x}$ is the sample mean

Let us consider some examples.

Example 1

Find the mean score of 10 students in a midterm exam in a class if their scores are:

25	27	30	23	16	27	29	14	20	28

Solution: The variable here is the scores of the students in the class, if X represents the variable then the values of X are

$$x_1 = 25, x_2 = 27, x_3 = 30, x_4 = 23, x_5 = 16, x_6 = 27, x_7 = 29,$$

$$x_8 = 14, x_9 = 20, \text{and} x_{10} = 28$$

The sum of all scores is

$$\begin{aligned}\sum x &= x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} \\ &= 25 + 27 + 30 + 23 + 16 + 27 + 29 + 14 + 20 + 28 \\ &= 239\end{aligned}$$

Since the given data set includes all scores of the students, it represents the population. Hence, $N = 10$. We have

$$\mu = \frac{\sum x}{N} = \frac{239}{10} = 23.9$$

Example 2

According to example 1, if we take a sample of 4 students from the class and find their scores to be: 23, 27, 16, and 29. Find the mean of this scores.

Solution:

Since the given data set includes some of the scores of the students, it represents the sample. Hence, $n = 4$. We have

$$\bar{x} = \frac{\sum x}{n} = \frac{23 + 27 + 16 + 29}{4} = \frac{95}{4} = 23.75$$

2. The Mean for Grouped Data:

According to definition 3.1.1, calculation of the mean for grouped data set depends on the sum of all values of the interest data set and their numbers. But when the data sets are in type of ungrouped, then it is impossible to find the sum of all values or their numbers, in such case an approximate to the sum of all values is used to calculate the mean. This approximation is illustrated by the following definition

Definition 3.1.2

Mean for population data: $\mu = \dfrac{\sum x_m f_m}{N}$

Mean for sample data: $\overline{x} = \dfrac{\sum x_m f_m}{n}$

Where x_m is the midpoint and f_m is the frequency of the class m.

Let us consider examples

Example 3

Find the population mean of the following frequency distribution.

Class	Frequency f_m
0 – 4	4
5 – 9	9
10 – 14	6
15 – 19	4
20 - 24	2

Solution: We construct the following table:

Class	Frequency f_m	x_m	$x_m f_m$
0 – 4	4	2	8
5 – 9	9	7	63
10 – 14	6	12	72
15 – 19	4	17	68
20 - 24	2	22	44
Total	**25**		**255**

Applying the formula for population mean, we get

$$\mu = \frac{\sum x_m f_m}{N} = \frac{255}{25} = 10.2$$

The following example is concerned with calculation the sample mean

Example 4

Find the sample mean for the following frequency distribution.

Class	Frequency f_m
5 – 9	4
10 – 14	8
15 – 19	15
20 – 24	10
25 - 29	7

Solution: Since the data set includes only a sample of 40 days, it represents a sample. So we construct the following table

Class	**Frequency** f_m	x_m	$x_m f_m$
5 – 9	4	7	28
10 – 14	8	12	96
15 – 19	15	17	255
20 – 24	10	22	220
25 - 29	7	27	189
Total	**44**		**788**

Applying the formula for sample mean, we get

$$\bar{x} = \frac{\sum x_m f_m}{n} = \frac{788}{44} = 17.9$$

B Weighted Mean

Assume that we have two classes A and B in PY, if the number of students in class A is 20 students, and in B is 30 students, moreover suppose that the grades in each class on stat 140 test are as follows:
In class A the grades are:
62, 67, 71, 74, 76, 77, 78, 79, 79, 80, 80, 81, 81, 82, 83, 84, 86, 89, 93, 98
In class B the grades are:
81, 82, 83, 84, 85, 86, 87, 87, 88, 88, 89, 89, 89, 90, 90, 90, 90, 91, 91, 91, 92, 92, 93, 93, 4, 95, 96, 97, 98, 99

Note that, the means for the two classes are 80 and 90 in class A and class B respectively. The mean of the two class means is 85 that is the mean of 80 and 90. However, the number of students on the two classes is different and hence the value 85 does not take in consideration the difference; therefore the value of 85 does not reflect the average student grades. In such case the average student grade in the two classes can be obtained by averaging all the grades, without regard to classes, that is by adding all grades in the two classes and dividing by the total number of students

$$\bar{x} = \frac{4300}{50} = 86.$$

Another way can be used to find the average which is by weighting means

Definition 3.1.3

The weighted mean of a variable can be found by multiplying each value by its corresponding weight and dividing the sum of the products by the sum of the weights

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum wx}{\sum w}$$

Where $w_1, w_2, \ldots, w_n$ are the weights and $x_1, x_2, \ldots, x_n$ are the values.

Let us consider an example

Remark

The weighted mean makes it possible to find the average for each group in the case where only the group's means and the number of observations in each group are known.

Example 5

A student received an A in English course (3 credits), a C in Math 130 (2 credits), a B in psychology (3 credits), and a D in computer science (4 credits). Assuming that $A = 4$grade points, $B = 3$ grade points, $C = 2$ grade points and $D = 1$ grade point. Find the mean of the grades of the student.

Solution: We construct the following table:

Course	Credits (w)	Grade (x)
English	3	$A = 4$ Points
Math 130	2	$C = 2$ Points
Psychology	3	$B = 3$ Points
Computers	4	$D = 1$ Points

Applying the formula for the weighted mean, we get

$$\bar{x} = \frac{\sum wx}{\sum w} = \frac{3(4) + 2(2) + 3(3) + 4(1)}{3 + 2 + 3 + 4}$$

$$= \frac{29}{12} = 2.42 \text{ Points}$$

C The Midrange

An approximate of the middle point for a data set is called midrange. Such approximation can be found as it is illustrated in the following definition

Definition 3.1.4

The midrange (MR) is defined as the sum of the lowest and highest values in the data set divided by 2.

$$MR = \frac{\text{Lowest value} + \text{Highest value}}{2}$$

Let us consider an example

Example 6

Find the midrange (MR) for the following data:

11, 13, 20, 30, 9, 4, 15

Solution: The lowest value is 4, and the highest value is 30, then

$$MR = \frac{4 + 30}{2} = \frac{34}{2} = 17$$

Note that, this measure (MR) is weak as a measure of central tendency since it is depends only on two values among of all values in the data set.

D The Median

The following definition of a central tendency helps us to overcome the problem of extreme values and the loss of some of the data (but under specific conditions), that is, it provides a partial solution to the problem of the loss of some data.

1. The Median for Ungrouped Data

A measure of central tendency that represents the middle term of a ranked data set is called the **median**

Definition 3.1.5

The median is the value of the middle term in a data set that has been ranked in increasing or decreasing order.

Note that, to find the median of a given data we need the following three steps

1. Rank the given data sets (in increasing or decreasing order)
2. Find the middle term for the ranked data set that obtained in step 1.
3. The value of this term represents the median.

In general form, calculating the median depends on the number of observations (even or odd) in the data set, therefore applying the above steps requires a general formula. The formula for calculating the median for the two cases (even and odd) follows:

Remark
According to the definition of the median, it is divides the ranked data set into two equal parts.

The median of the ranked data $x_1, x_2, \ldots, x_n$ is given by

$$\text{Median} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{if } n \text{ is even} \end{cases}$$

Let us consider an examples

Example 7

Find the median for the data set:

312, 257, 421, 289, 526, 374, 497

Solution: First, the data set after we have ranked in increasing order is:

x_1	x_2	x_3	x_4	x_5	x_6	x_7
257	289	312	374	421	497	526

Median=374

Since there are 7 values in this data set, so the fourth term $\left(\frac{7+1}{2} = 4\right)$ in the ranked data is the median. Therefore the median is

$$\text{median} = x_{\frac{n+1}{2}} = x_{\frac{7+1}{2}} = x_4 = 374$$

Example 8

Find the median for the data set:

8, 12, 7, 17, 14, 45, 10, 13, 17, 13, 9, 11

Solution: First, we rank the data in increasing order:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
7	8	9	10	11	**12**	**13**	13	14	17	17	45

Median=12.5

Since there are 12 values in this data set, the median is given by the average of the two middle values whose ranks are $\left(\frac{12}{2}=6\right)$ and 6+1=7. Therefore, the median is

$$\text{median} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{\frac{12}{2}} + x_{\frac{12}{2}+1}}{2} = \frac{x_6 + x_7}{2}$$

$$= \frac{12+13}{2} = 12.5$$

1. The Median for Grouped Data

Suppose that we have a frequency distribution table with k classes, then one calculate the median of this grouped data by the following relation:

$$\text{median} = L_{med} + \frac{\frac{\sum_m f_m}{2} - \left(F_{med} - f_{med}\right)}{f_{med}} \times C_{med}$$

Where

L_{med} is the lower boundary of the median class, note that the median class is the first class have a ascending cumulative frequency greater or equal to half of the total frequencies,

f_{med} is the frequency the median class,

F_{med} is the ascending cumulative frequency of the median class,

C_{med} is the class width of the median class, and

$\sum_m f_m$ is the sum of total frequencies.

Let us consider the following example

Example 9

Let be a frequency distribution table given as follow:

m	Class Boundaries	Midpoint x_m	Frequency f_m	Ascending Cumulative Frequency F_{med}
1	40 – 50	45	2	2
2	50 – 60	55	5	7
3	60 – 70	65	10	17
4	70 – 80	75	15	32
5	80 – 90	85	9	41
6	90 – 100	95	5	46
Sum			**46**	

Solution: To calculate the median of this grouped data we must determine the median class. We note that the median class is the fourth class because it is the first class have a ascending cumulative frequency greater or equal to half of the total frequencies. So we have:

k=6, $L_{med} = 70,\ f_{med} = 15,\ F_{med} = 32,\ C_{med} = 10$ and $\sum_m f_m = 46$

Therefore the median for the given grouped data is:

$$\text{median} = L_{med} + \frac{\frac{\sum_m f_m}{2} - \left(F_{med} - f_{med}\right)}{f_{med}} \times C_{med}$$

$$= 70 + \frac{\frac{46}{2} - \left(32 - 15\right)}{15} \times 10 = 70 + 4 = 74$$

E The Mode

The **mode** is another measure of central tendency and it is known as the most common value in a data set. The definition of the mode is given below

Definition 3.1.6

The mode is the value that occurs most often in a data set.

Let us consider the case of ungrouped data

1. Mode for Ungrouped Data

For a given ungrouped data set there are several cases of mode, a data set may have none or have one mode and may have more than one mode. For the three cases we define the following:

1. Data set with none mode: In such data set each value occurring only once.

2. Data set with one mode: In such data set only one value occurring with the highest frequency. The data set in this case is called **unimodal**.

3. Data set with two modes: In such data set two values that occur with the same (highest) frequency. The distribution, in this case, is said to be **bimodal**.

4. Data set with more than two modes: In such data set more than two values occurs with the same (highest) frequency, then the data set contains more than two modes and it is said to be **multimodal**.

Let use consider some examples

Example 10

Find the mode for the given data set:

76, 150, 95, 750, 124, 985, 87, 490

Solution: Because each value in this data set occurs only once, this data set contains no mode.

Example 11

Find the mode for the given data set:

77, 82, 74, 81, 79, 84, 74, 78

Solution: Here the value 74 occurs twice, that is with the highest frequency, so 74 is the mode.

Example 12

Find the mode for the given data set:

7, 8, 9, 10, 11, 12, 13, 13, 14, 17, 17, 45

Solution: Since each of the two values, 13 and 17 occurs twice and each of the remaining values occurs only once. Therefore, this data set has two modes: 13 and 17.

Remark

The distribution of a data set has one choice of the following:

(i) None mode.

(ii) Unimodal distribution.

(iii) Bimodal distribution

(iv) Multimodal distribution.

Example 13

Find the mode for the given data set:

22, 19, 21, 19, 27, 21, 29, 22, 19, 25, 21, 22, 25

Solution: Since each of the three values, 19 (occur three times), 21 (occur three times), and 22 (occur three times) occurs with a highest frequency in their neighborhoods, therefore, each of these is a mode, that is the modes for this data set are: 19, 21, and 22.

2. Mode for Grouped Data

Suppose that we have a frequency distribution table with k classes, then one calculate the mode of this grouped data by the following relation:

$$\text{mode} = L_{\text{mod}} + \frac{d_1}{d_1 + d_2} \times C_{\text{mod}}$$

where

L_{mod} is the lower boundary of the mode class, note that the mode class is the class that has a frequency greater or equal to the frequency of previous and subsequent classes. Should not be the first or the last class,

d_1 is the deference between the frequency of the mode class and the frequency of the previous class,

d_2 is the deference between the frequency of the mode class and the frequency of the next class, and

C_{mod} is the class width of the mode class.

Example 14

Refer to example 9 to calculate the mode for the given frequency distribution table:

Solution: To calculate the mode of those grouped data we must determine the mode class. We not that the mode class is the fourth class because it has a frequency greater than the frequency of her former and subsequent class. So we have

$L_{\text{mod}} = 70$, $d_1 = 15 - 10 = 5$, $d_2 = 15 - 9 = 6$ and $C_{\text{mod}} = 10$

Therefore the mode for the given grouped data is

$$\text{mode} = L_{mod} + \frac{d_1}{d_1 + d_2} \times C_{mod}$$

$$= 70 + \frac{5}{6+5} \times 10 = 70 + 4.545$$

$$= 74.545$$

Example 15

For the following raw data

22, 21, 23, 24, 25, 24, 23, 22, 5, 2, 4, 3, 1, 2, 4, 16, 17, 18, 17, 18, 19, 17, 18, 16, 20, 20, 18, 10, 9, 7, 8, 6, 9, 8, 6, 9, 11, 13, 15, 12, 14, 14, 12, 13, 15, 11, 11, 14, 13, 12

a) Calculate the mean, median and mode.

b) Molding this data in a frequency distribution table with class limit width equal to 4 (class boundaries width equal to 5).

c) Calculate the mean, median and mode of data for the frequency distribution table.

d) Compare the values of mean, median and mode for ungrouped data and th grouped data. What do you notice, and why did this happen?

Solution: For (a),

1) To calculate the mean we have

$$\mu = \frac{\sum x}{N} = \frac{22 + 21 + \cdots + 13 + 12}{50} = 13.42$$

If the data for the population

$$\overline{x} = \frac{\sum x}{n} = \frac{22 + 21 + \cdots + 13 + 12}{50} = 13.42$$

If the data for the sample

Note that the mean value does not change, whether the data to the population or to sample.

2) To calculate the median we must first arrange the data in ascending (or descending), where we find after arranged data in ascending order as follows:

1, 2, 2, 3, 4, 4, 5, 6, 6, 7, 8, 8, 9, 9, 9, 10, 11, 11, 11, 12, 12, 12, 13, 13, **13, 14,** 14, 14, 15, 15, 16, 16, 17, 17, 17, 18, 18, 18, 18, 19, 20, 20, 21, 22, 22, 23, 23, 24, 24, 25

Now, as an even number of data (equal to 50), the median value can be calculated as follows:

$$\text{median} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{\frac{50}{2}} + x_{\frac{50}{2}+1}}{2} = \frac{x_{25} + x_{26}}{2}$$

$$= \frac{13+14}{2} = 13.5$$

3) To determine the value of the mode we know that the mode is the most frequent value, and therefore the mode is 18, because it is the only value that four times and the rest of the values have frequencies less than four.

For b) With molding given raw data in a distribution frequency table with class limit width equal 4 (class boundaries width equal 5). We getting of the following table:

m	Class Limit	Class Boundaries	Midpoint x_m	Frequency f_m	Ascending Cumulative Frequency F_{med}
1	1 – 5	0.5 - 5.5	3	7	7
2	6 – 10	5.5 - 10.5	8	9	16
3	11 – 15	10.5 - 15.5	13	14	30
4	16 – 20	15.5 - 20.5	18	12	42
5	21 – 25	20.5 - 25.5	23	8	50
Sum				**50**	

For c)

1) The mean of grouped data we have

$$\bar{x} = \frac{\sum x_m f_m}{\sum f_m} = \frac{\sum x_m f_m}{n}$$

$$= \frac{3(7) + 8(9) + 13(14) + 18(12) + 23(8)}{50}$$

$$= \frac{675}{50} = 13.5$$

For the population data is the same

2) The median of the grouped data we must determine the median class. We not that the median class is the third class because it is the first class have a ascending cumulative frequency greater or equal to half of the total frequencies. So we have

k=5, $L_{med} = 10.5$, $f_{med} = 14$, $F_{med} = 30$, $C_{med} = 5$ and $\sum_m f_m = 50$

Therefore the median for the given grouped data can be calculated as follows:

$$\text{median} = L_{med} + \frac{\frac{\sum_m f_m}{2} - \left(F_{med} - f_{med}\right)}{f_{med}} \times C_{med}$$

$$= 10.5 + \frac{\frac{50}{2} - (30 - 14)}{14} \times 5 = 10.5 + 3.21 = 13.71$$

3) The mode of the grouped data we must determine the mode class. We not that the mode class is the third class because it has a frequency greater than the frequency of her former and subsequent class. So we have

$L_{mod} = 10.5$, $d_1 = 14 - 9 = 5$, $d_2 = 14 - 12 = 2$ and $C_{mod} = 5$

Therefore the mode for the given grouped data can be calculated as follows:

$$\text{mode} = L_{mod} + \frac{d_1}{d_1 + d_2} \times C_{mod}$$

$$= 10.5 + \frac{5}{5+2} \times 5 = 10.5 + 3.57$$

$$= 14.07$$

For d)

We note that the values of mean, median and mode for grouped data are deferent from the ungrouped data. The reason for this is because the classes in the classified data to be represented in their midpoints.

Exercises 3.1

1 For each of the following, find the mean, median and the mode

a 13, 11, 10, 5, 4, 2, 10, 6

b 2, 0, 0, 4, 6, 10, 0

c -12, -10, -20, -24, -14

d 3, 0, -10, 20, 7, 3, 7, 12, 1, -4, 7

2 The students in a math class took the Scholastic Aptitude Test. Their scors are shown below. Find the mean score.

640	618	345	349	574
348	356	581	470	482

3 Last year, nine employees of an electronic company retired. Their ages at retirement are listed below. Find the mean retirement age.

50	62	61	52	62
58	65	52	55	

4 The mean age of six persons is 49 years. The ages of five of these six persons are: 55, 39, 44, 51 and 45 years respectively. Find the age of the sixth person.

5 Suppose a sample of 10 statistics books gave a mean price of SAR 140 and a sample of 8 mathematics books gave a mean price of SAR 160. Find the combined (weighted) mean.

6 The mean monthly income for five families was SAR 5000. What is the total monthly income of these five families.

7 Exam scores for 100 randomly selected college students has the frequency distribution was shown in the table below

Class	Frequency f_m
90 – 98	5
99 – 107	20
108 – 116	40
117 – 125	30
126 - 134	5

Find the mean, median and mode of the exam scores above.

8 Find the mean, median and mode for the following frequency distributions.

a

Class	Frequency f_m
7.5 – 12.5	3
12.5 – 17.5	5
17.5 – 22.5	15
22.5 – 27.5	5
27.5 – 32.5	2

b

Class	Frequency f_m
10 – 20	1
21 – 31	9
32 – 42	14
43 – 53	8
54 – 64	10
65 – 75	3

9 Find the midrange for the data set:

a 12, 10, 3, 11, 30, 41

b 5, -3, 2, 21, 19, 20

10 Suppose an instructor gives two exams and a final, assigning the final exam a weight twice that of each of the other exams. Find the weighted mean for a student who scores 73 and 67 on the first two exams and 85 on the final exam.
(**Hint**: $x_1 = 73,\ x_2 = 67,\ x_3 = 85,\ w_1 = w_2 = 1$and $w_3 = 2$)

11 The table below shows the mean and number of students in three classes in an exam

Class	Mean	Number
A	75	25
B	65	20
C	70	30

Find the weighted mean for all students in the three classes above.

12 Find the median for the given sample data:

a 11, 13, 18, 21, 30, 30,49

b 15, 19, 40, 55, 70, 72, 85

c 75, 62, 217, 159, 299, 230, 230

13 Find the mode(s) for the given sample data a -20, -23, -46, -23, -49, -23, -49, -50, -48.

b 20, 21, 46, 21, 49, 21, 49

c 92, 57, 32, 57, 29, 21, 49

14 Last year, ten employees of an electronics company retired. Their ages at retirement are listed below. Find the mode(s).

10	12	11	12	12
18	18	12	18	18

15 The mean of the data 18,18,10,18,20 is

A)16 B)16.8 C)18 D)21

16 The grocery expenses for six families were $55.72, $55.08, $76.11, $54.18, $63.56, and $85.72 Compute the mean grocery bill. Round your answer to the nearest cent.

A) $97.59 B) $78.07 C) $65.06 D) $66.07

17 The median of the data 11,13,18, 21, 30, 30, 49 is

A) 18 B) 25.5 C) 30 D) 21

18 The mode(s) for the data -20,-23,-46,-23,-49,-23,-49 is

A) -33.3 B) -49 C) -23 D) -46

19 The range for the data 26 40 17 42 59

A) 59 B) 14 C) 17 D) 42

20 The sample mean for the following frequency distribution is

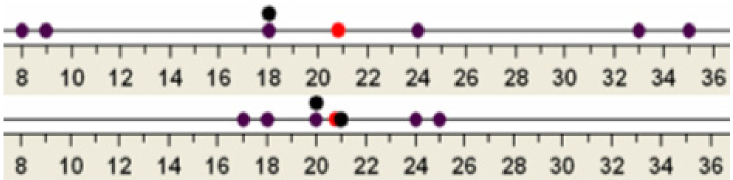
Class	Frequency f_m
5 – 9	2
10 – 14	3
15 – 19	2
20 – 24	1
25 - 29	2

A) 160 B) 16 C) 16.2 D) 16.3

3.2 Measures of Variation

In section 3.1, we have discussed the measures of central tendency, such as the mean, median, and mode, these measures give us a picture of the distribution of a data set, but this picture is unclearly. For example: let be the following two data sets:

X	8	9	18	35	33	24	18
Y	17	18	20	25	24	21	20



We find that the average of data in both data sets is 20.7 but the variation of values varies from one to another. So we note that for two data sets with the same average, the variation among the values of observations for the first data set may be much larger or smaller than for the second data set.
In fact there are many measures of variation and the simplest of it is the following scale:

A Range

In simple words, the **range** for a data set is depends on two values (the smallest and the largest values) among all values in such data set. The definition of the range is given below

Definitions 3.2.1

1. The range for ungrouped data is defined by

$$\text{Range} = \text{Largest value} - \text{Smallest value}$$

2. The range of grouped data in k classes we use the following relation:

$$\text{Range} = x_k - x_1$$

Where x_1 and x_k are the midpoints of the first and the last class respectively.

Example 1

Find the range for the data set:

40, 10, 20, 30, 35, 40, 50, 60

Solution: The largest value is 60, and the smallest value is 10. Therefore

$$\text{Range} = \text{Largest value} - \text{Smallest value}$$
$$= 60 - 10 = 50$$

Example 2

Use the following distribution to find the range

Staff	Salary
Owner	SAR 400,000
Manager	SAR 160,000
Sales representative	SAR 120,000
Workers 1	SAR 100,000
Worker 2	SAR 60,000
Worker 3	SAR 72,000

Solution: The largest value is 400,000, and the smallest value is 60,000, then

$$\text{Range} = \text{Largest value} - \text{Smallest value}$$
$$= 400{,}000 - 60{,}000 = \text{SAR } 340{,}000$$

Remark

The range is the simplest measure of variation, and can be misleading because it ignores the way in which data are distributed. And more the range is sensitive to extreme values. With another words the range is very sensitive to the smallest and largest data values.

Example 3

Consider the grouped data in example 9 of previous section, then find the range.

Solution: Refer to such data, we find that $x_1 = 45$ and $x_k = 95$ and then the range of this grouped data is

$$\text{Range} = x_k - x_1 = 95 - 45 = 50$$

The following scale provides us with a better measure of the variation:

B Mean Deviation

Another measure of variation is called **mean deviation**; it is the mean of the distances between each value and the mean

The mean deviation is found by the formula:

$$\text{Mean deviation} = \frac{\sum |x - \overline{x}|}{n}$$

Where x is a value, $\overline{x}$ is the mean, and n is the number of values

Remark

The algebraic sum of the deviations $\sum (x - \overline{x})$ always equal to zero, e.g. $\Sigma (x - \overline{x}) = 0$. Therefore we take the absolute value for each deviation to compute the mean deviation.

Let us consider an examples

Example 4

Find the mean deviation for the sample data set: 5, 7, 11, 13.

Solution:

Step1: First we must find the sample mean as follows:

$$\overline{x} = \frac{5+7+11+13}{4} = \frac{36}{4} = 9$$

Step2: Use the value that obtained in step 1 to complete the following table

x	$x - \overline{x}$	$\lvert x - \overline{x} \rvert$
5	-4	4
7	-2	2
11	2	2
13	4	4
Sum	**0**	**12**

Appling the formula of the mean deviation we get

$$\text{Mean deviation} = \frac{\sum |x - \overline{x}|}{n} = \frac{12}{4} = 3$$

In order to calculate the mean deviation of grouped data in k classes we use the following relation:

$$\text{mean deviation} = \frac{1}{\sum f_m} \sum f_m \cdot |x_m - \overline{x}|$$

Where:

x_m is the midpoint of the m^{th} class,

f_m is the frequency of the m^{th} class, and

$\overline{x}$ is the mean of grouped data.

Example 5

Consider the grouped data in example 9 of previous section, then find the range.

Solution: Recall that the grouped data in example 9 is

m	Class Boundaries	Midpoint x_m	Frequency f_m
1	40 – 50	45	2
2	50 – 60	55	5
3	60 – 70	65	10
4	70 – 80	75	15
5	80 – 90	85	9
6	90 – 100	95	5
	Sum		**46**

It can be easy find the mean of this grouped data by applying the formula

$$\overline{x} = \frac{\sum x_m f_m}{\sum f_m} = 73.478$$

Now, we find the mean deviation of the given grouped data is:

$$\text{mean deviation} = \frac{1}{\sum_m f_m} \sum_m f_m \cdot \left| x_m - \overline{x} \right|$$

$$= \frac{2\left| 45 - 73.478 \right| + \cdots + 5\left| 95 - 73.478 \right|}{46}$$

$$= \frac{468.264}{46} = 10.18$$

It has been observed that if we take the square differences between values and its average, we will get a better measure of the dispersion, and this is what will do to the following scale.

C The Variance and Standard Deviation

A most used measure of variation is called **standard deviation** denoted by (σ for the population and S for the sample). The numerical value of this measure helps us how the values of the dataset corresponding to such measure are relatively closely around the mean. Without loss of generality, in case of a lower value of the standard deviation for a data set, we say that the values of such data set are spread over a relatively smaller range around the mean; conversely, in case of a larger value of the standard deviation for a data set, we say that the values of such data set are spread over a relatively smaller range around the mean. Another measure of variation is called the variance denoted by (σ^2 for the population and S^2 for the sample); such measure can be found by squared the standard deviation.

30 For Ungrouped Data

The population variance σ^2 and the sample variance s^2 and the standard deviations for an ungrouped data set are defined as follows:

Definition 3.2.3

The population variance σ^2 and the sample variance S^2 are the average of the squares of the distance (deviation) for each value from the population mean and the sample mean, selectively.

The formulas that are used to calculate the population variance and the population standard deviation are given by

$$\sigma^2 = \frac{\sum (x-\mu)^2}{N} \quad \text{and} \quad \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

Where x is individual value, μ is the population mean, and N is the population size

This relation is usually used when we have the mean value of population, whereas the formulas that are used to calculate the sample variance and the standard deviation are given by

$$S^2 = \frac{\sum (x-\overline{x})^2}{n-1} \quad \text{and} \quad S = \sqrt{S^2} = \sqrt{\frac{\sum (x-\overline{x})^2}{n-1}}$$

Where x is individual value, $\overline{x}$ is the sample mean, and n sample size

This relation is usually used when we have the mean value of sample.

Moreover, another simplified formula can be used to calculate the population variance and the sample variance are given by

$$\sigma^2 = \frac{\sum x^2 - N\mu^2}{N} = \frac{\sum x^2}{N} - \mu^2 \text{ and } S^2 = \frac{n\sum x^2 - \left(\sum x\right)^2}{n(n-1)}$$

This relationship is usually used when we want to use the raw data directly.

Let us consider some examples

Example 6

Find the variance and the standard deviation for the given data set:

20, 30, 15, 20, 25, 10

Solution:

Step1: First we must find the population mean as follows:

$$\mu = \frac{20 + 30 + 15 + 20 + 25 + 10}{6} = \frac{120}{6} = 20$$

Step2: Subtract the mean from each value.

(20-20=0), (30-20=10), (15-20=-5), (20-20=0), (25-20=5), and (10-20=-10).

Step3: Square the results in step2 and fill the following table:

x	$x-\mu$	$(x-\mu)^2$
20	0	0
30	10	100
15	-5	25
20	0	0
25	5	25
10	-10	100
Sum	**0**	**250**

Applying the formula of the population variance, we get

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} = \frac{250}{6} = 41.66$$

Applying the formula of the population standard deviation, we get

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}} = \sqrt{41.66} = 6.5$$

Example 7

Apply the two formulas of the sample variance and the standard deviation for the given sample:

3, 5, 7, 9, 11

Solution:

Step1: First we must find the sample mean as follows:

$$\bar{x} = \frac{3+5+7+9+11}{5} = \frac{35}{5} = 7$$

Step2: Use the value that obtained in step 1 to fill the table below:

x	$x-\bar{x}$	$(x-\bar{x})^2$
3	-4	16
5	-2	4
7	0	0
9	2	4
11	4	16
Sum	**0**	**40**

Applying the formulas of the sample variance and the standard deviation, we get

$$S^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{40}{5-1} = \frac{40}{4} = 10$$

$$S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{10} \simeq 3.2$$

According to the simplified formulas, we must construct the following table

	x	x^2
	3	9
	5	25
	7	49
	9	81
	11	121
Sum	**35**	**285**

Applying the simplified formula for the sample variance, we get

Remark

Referring to the formulas of the population variance or the sample variance, then the formulas includes the quantities $(x-\mu)$ and $(x-\bar{x})$ which they are called the deviation of the x value and the mean.

$$S^2 = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)} = \frac{5(285) - (35)^2}{5(5-1)}$$

$$= \frac{1425 - 1225}{20} = \frac{200}{20} = 10$$

Applying the simplified formula for the sample standard deviation, we get

$$S = \sqrt{\frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)}} = \sqrt{10} \simeq 3.2$$

2. For Grouped Data

For the grouped data or frequency distribution, it is necessarily to give basic formulas that used to calculate the population and sample variance and their standard deviations. Basic formulas that are used are given as follows

The formulas for the population variance and the population standard deviation are given by

$$\sigma^2 = \frac{\sum f_m (x_m - \mu)^2}{N} \quad \text{and} \quad \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f_m (x_m - \mu)^2}{N}}$$

Where x_m is the midpoint of class m, f_m is the frequency of class m, and μ is the population mean

This relation is usually used when we have the mean value of population, whereas the formulas can be used to calculate the sample variance and the standard deviation are given by

$$S^2 = \frac{\sum f_m (x - \overline{x})^2}{n-1} \quad \text{and} \quad S = \sqrt{S^2} = \sqrt{\frac{\sum f_m (x - \overline{x})^2}{n-1}}$$

Where x_m is the midpoint of class m, f_m is the frequency of class m, and $\overline{x}$ is the sample mean

This relation is usually used when we have the mean value of sample. Moreover, another simplified formulas can be used to calculate the population variance and the sample variance are given by

$$\sigma^2 = \frac{\sum (x_m)^2 f_m - \frac{\left(\sum x_m f_m\right)^2}{N}}{N} \quad \text{and} \quad S^2 = \frac{n\sum f_m \cdot (x_m)^2 - \left(\sum f_m \cdot x_m\right)^2}{n(n-1)}$$

This relationship is usually used when we want to use the raw data directly

This relation is usually used when we have the mean value of sample.
Let us consider an example

Example 8

Find the variance and the standard deviation for the following frequency distribution of a sample:

Class	Frequency f_m
5 – 9	2
10 – 14	4
15 – 19	7
20 – 24	3
25 - 29	1
30 - 34	**3**
Total	**20**

Solution: First we must construct the following table:

Class	Frequency f_m	x_m	$x_m f_m$	$(x_m)^2 f_m$
5 – 9	2	7	14	98
10 – 14	4	12	48	576
15 – 19	7	17	119	2023
20 – 24	3	22	66	1452
25 - 29	1	27	27	729
30 - 34	3	32	96	3072
Sum	**20**		**370**	**7950**

Then we have the mean of frequency distribution equal to:

$$\bar{x} = \frac{\sum_m x_m f_m}{\sum_m f_m} = \frac{370}{20} = 18.5$$

And the variance equal to:

$$S^2 = \frac{\sum_m f_m (x_m - \bar{x})^2}{\left(\sum_m f_m\right) - 1} = \frac{\sum_m f_m (x_m - \bar{x})^2}{n-1}$$

$$= \frac{264.5 + 169 + 15.75 + 36.75 + 72.25 + 546.75}{19} = \frac{1105}{19} = 58.158$$

Therefore we get the standard deviation equal to:

$$S = \sqrt{S^2} = \sqrt{58.158} \simeq 7.626$$

Or that we are using the raw data directly, then we have:

$$S^2 = \frac{n\sum f_m \cdot (x_m)^2 - \left(\sum f_m \cdot x_m\right)^2}{n(n-1)}$$

$$= \frac{20(7950) - (370)^2}{20(19)} = \frac{159000 - 136900}{380}$$

$$= 58.158$$

Using the definition of the standard deviation, we get

$$S = \sqrt{S^2} = \sqrt{58.158} \simeq 7.626$$

D The Coefficient of Variation

A coefficient of variation (CV) is one of well known measures that used to compare the variability of two different data sets that have different units of measurement. Moreover, one disadvantage of the standard deviation that its being a measure of absolute variability and not of relative variability. The following definition is displayed such measure and how can be calculated.

Definition 3.2.4

The coefficient of variation, denoted by (CV), expresses standard deviation as a percentage of the mean and is computed as follows

For population data $CV = \frac{\sigma}{\mu} \cdot 100\%$; $\mu \neq 0$

For sample data $CV = \frac{S}{\bar{x}} \cdot 100\%$; $\bar{x} \neq 0$

Note that the value of the coefficient of variation is an abstract value (without unit of measurement)

Let us consider some examples

Example 9

Find the coefficient of variation for the sample data:

10, 20, 30, 40, 50

Solution:

Step1: First, it must find the sample mean

$$\bar{x} = \frac{10 + 20 + 30 + 40 + 50}{5} = \frac{150}{5} = 30$$

Step2: Using the value obtained in step 1 to find the sample standard deviation by construction the table:

x	$x-\overline{x}$	$(x-\overline{x})^2$
10	-20	400
20	-10	100
30	0	0
40	10	100
50	20	400
Sum	**0**	**1000**

Using the definition of the sample standard deviation, we get

$$S^2 = \frac{\sum (x-\overline{x})^2}{n-1} = \frac{1000}{5-1} = \frac{1000}{4} = 250$$

$$S = \sqrt{\frac{\sum (x-\overline{x})^2}{n-1}} = \sqrt{250} = 5\sqrt{10} = 15.8$$

Using the definition of the coefficient of variation, we get

$$CV = \frac{S}{\overline{x}} \cdot 100\%$$

$$= \frac{15.8}{30} \cdot 100\% = 52.67\%$$

Example 10

Assume that we have two samples of scores A and B, if the mean of scores for class A is 85 and the standard deviation is 5. The mean of scores for class B is 90 and the standard deviation is 8. Compare the variation of the two classes.

Solution: To compare the variability of the two classes, we must compare their coefficient of variations, doing so we get:

For the class A the coefficient of variation is

$$CV = \frac{S}{\overline{x}} \cdot 100\% = \frac{5}{85} \cdot 100\% = 5.85\%$$

For the class B the coefficient of variation is

$$CV = \frac{S}{\overline{x}} \cdot 100\% = \frac{8}{90} \cdot 100\% = 8.88\%$$

Since the coefficient of variation for the class B is larger than the coefficient of variation for the class A, then the scores in class B are more variable than the scores in the other class B.

Exercises 3.2

1 Find the range for each data set of the following

a 12, 4, -10, 8, 8, -13

b 4, -7, 1, 0, -9, 16, 9, 8

c 20, 5, 31, 40, 19, 61, 18

d 1, 3, 7, 11, 19, 5

2 Find the mean deviation for the following data set

a 2, 4, 6, 8, 10

b 10, 20, 30, 40, 50

c -5, -10, -15, -20, -25

d 1, -1, 2, -2, 5

3 Find the variance and standard deviation for the following data sets

a 1, 3, 5, 7, 9, 11

b 2, 4, 6, 8, 10, 12

c 3, 6, 9, 12, 15

d -2, -3, -4, -5, -6

4 Following are the temperatures (in degrees Celecius) observed during 7 days in a city:

11, 23, 6, 14, -2, 16, 19

Find the range, variance, and the standard deviation.

5 The following data give the numbers of hours studying by 10 randomly selected college students during the past week:

5, 7, 14, 0, 7, 9, 4, 10, 0, 8

Calculate the range, variance, and the standard deviation.

6 The SAT scores of 100 students have a mean of 975 and a standard deviation of 105. The GPAs of the same 100 students have a mean of 3.16 and a standard deviation of 0.22. Is the relative variation in SAT scores larger or smaller than in GPAs.

7 Using the population formula, calculate the variance, and standard deviation for the following frequency distribution:

Class	4 - 5	7 - 8	10 – 11	13 - 14	16 – 17
f	5	9	14	7	5

8 Using the sample formula, compute the variance, and standard deviation for the following frequency distribution:

Class	1 - 5	6 - 10	11 - 15	16 - 20	21 - 25	26 - 30
f	5	13	10	8	8	6

9 The following data are ages (in years) of 10 students

20, 20, 20, 20, 20, 20, 20, 20, 20, 20

Find the standard deviation. Comment on your answer.

10 Consider the following two data sets:

Data set I: 13, 25, 27, 15

Data set II: 10, 15, 30, 25

Find the mean and the standard deviations for each data set of the two data sets above. Comment on your answer.

11 Fatima is currently taking college economics. The instructor often gives quizzes. On the past five quizzes, Fatima got the following scores

5, 19, 2, 13, 10

Find the range for her scores.

12 The owner of a small manufacturing plant employs six people. As part of their personal file, she asked each one to record to the nearest one-tenth of a mile the distance they travel one way from home to work. The six distances are listed below.

2.3, 5.5, 1.1, 4.3, 6.4, 3.5

Find the range for these data.

13 A class of sixth grade students kept accurate records on the amount of time they spent playing video games during a one-week period. The times (in hours) are listed below

26.7	14.7	8.3	12.9	15.1
28.7	23.0	23.6	14.3	11.0

Find the range for these data.

14 The test scores of 19 students are listed below

91	99	86	54	72
85	97	91	90	66
82	83	78	88	77
80	92	94	98	

Find the range for these data.

15 Find the sample standard deviation for the given data. Round your final answer to one more decimal place than that used for the observations

a 15, 42, 53, 7, 9, 12, 14, 28, 47

b 2, 6, 15, 9, 11, 22, 1, 4, 8, 19

16 The manager of an electrical supply store measured the diameters of the rolls of wire in the inventory. The diameters of the rolls (in m) are listed below

0.299, 0.173, 0.227, 0.177, 0.634, 0.621, 0.127

Find the sample standard deviation for the given data.

17 A company had 80 employees whose salaries are summarized in the frequency distribution below. Find the standard deviation and coefficient of variation.

Salary (Dollars)	Employees
5,001 – 10,000	12
10,001 – 15,000	13
15,001 – 20,000	14
20,001 – 25,000	19
25,001 – 30,000	**22**
Total	**80**

18 The sample standard deviation for the following data is

22, 29, 21, 24, 27, 28, 25, 36

A) 4.2 B) 2.8 C) 1.6 D) 4.8

19 The test scores of 40 students are summarized in the frequency distribution below. Find the standard deviation.

Salary (Dollars)	Employees
50 – 60	5
60 – 70	9
70– 80	10
80 – 90	8
90 – 100	**8**

Any of the following values (A, B, C or D) is the standard deviation of given data ?

A) S=12.5 B) S=11.9 C) S=13.2 D) S=13.9

20 Any of the following values (A, B, C or D) is the coefficient of variation ($C.V$) for a sample data with mean 20 and variance 9 ?

A) 15% B) 45% C) 1.6% D) 4%

3.3 Measures of Position

For a population or a sample data sets, the measures of position are a tool that used to determine the position of a single value in relation to other values. In this section we will discuss the standard Scores, quartiles, percentiles, and percentile rank.

A Standard Scores

A first measure of position is called standard score, to understand the concepts of the standard score, suppose that a student scored is 90 on a driving test, and 45 on an English exam. Direct comparison of raw scores is impossible, since the exams might not be equivalent in terms of number of questions, value of each question, and so on. However, a comparison of a relative standard similar to both can be made. This comparison uses the mean and standard deviation and it is called a standard score (or z- score).

Definition 3.3.1

A standard score (or z-score) for a value is obtained by the formula:

$$z = \frac{\text{value - mean}}{\text{standard deviation}}$$

For population: $z = \dfrac{x - \mu}{\sigma}$ and for sample: $z = \dfrac{x - \overline{x}}{s}$

Let us consider example

Example 1

A student scored 65 on a mathematics test that had a mean of 50 and standard deviation of 10; he scored 30 on a statistics test with a mean of 25 and a standard deviation of 5. Compare his relative position on the two tests.

Solution: To compare the relative position on the two tests, we must find the score for the two score of the students.

For mathematics test, the z-score is:

$$z = \frac{65 - 50}{10} = 1.5$$

For statistics test, the z-score is:

$$z = \frac{30 - 25}{5} = 1$$

Since z-score for mathematics is larger, his relative position in mathematics is higher than his relative position on statistics.

B Quartiles and Interquartile Range

Any data set can be divided into four equal parts by using a summary measure called quartiles. There are three quartiles that used to dived the data set which is denoted by (Q_i) for i=1,2,3. The following definition is illustrated the meaning of the quartiles.

Definition 3.3.2

Quartiles are three summary measures that divide a ranked data set into four equal parts. The second quartile is the same as the median, the first quartile is the middle term among the observations that are less than the median, and the third quartile is the value of the middle term among the observations that are greater than the median.

A measure of variation that depends on the first and the third quartiles is celled the interquartile range (IQR) and it is defined as follows

Definition 3.3.3

The difference between the third and the first quartiles gives the interquartile range, that is:

$$IQR = \text{Interquartile range} = Q_3 - Q_1$$

Let us consider an example

Example 2

For the given data set

8, 12, 7, 17, 14, 45, 10, 13, 17, 13, 9, 11

a. Find the values of the three quartiles.

b. Find the interquartile range.

Solution: First, we rank the given data in increasing order. Then we calculate the three quartiles as follows

$$\underbrace{7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12}_{\text{Values less than the median}} \quad \underbrace{13 \quad 13 \quad 14 \quad 17 \quad 17 \quad 45}_{\text{Values greater than the median}}$$

a. The values of the three quartiles are

$$Q_1 = \frac{9+10}{2} = 9.5, \quad Q_2 = \frac{12+13}{2} = 12.5, \quad Q_3 = \frac{14+17}{2} = 15.5$$

b. The interquartile range is

$$IQR = Q_3 - Q_1 = 15.5 - 9.5 = 6$$

Example 3

For the given data set

$$61, 24, 39, 51, 37, 59, 45$$

a. Find the values of the three quartiles.

b. Find the interquartile range.

Solution: : First, we rank the given data in increasing order. Then we calculate the three quartiles as follows

$$\underbrace{24\ \ 37\ \ 39}_{\substack{\text{Values less}\\ \text{than the median}}} \qquad \underbrace{45}_{\text{median }(Q_2)} \qquad \underbrace{51\ \ 59\ \ 61}_{\substack{\text{Values greater}\\ \text{than the median}}}$$

a. The values of the three quartiles are

$$Q_1 = 37, \quad Q_2 = 45, \quad Q_3 = 59$$

b. The interquartile range is

$$IQR = Q_3 - Q_1 = 59 - 37 = 22$$

Example 4

For the given data set

$$78, 64, 51, 45, 37, 59, 63, 24, 68$$

a. Find the values of the three quartiles.

b. Find the interquartile range.

Solution: First, we rank the given data in increasing order. Then we calculate the three quartiles as follows

$$\underbrace{24\ \ 37\ \ 45\ \ 51}_{\text{Values less than the median}} \quad \underbrace{59}_{\text{The median }(Q_2)} \quad \underbrace{63\ \ 64\ \ 68\ \ 78}_{\text{Values greater than the median}}$$

a. The values of the three quartiles are

$$Q_1 = \frac{37+45}{2} = 41, \quad Q_2 = 59, \quad Q_3 = \frac{64+68}{2} = 66$$

b. The interquartile range is

$$IQR = Q_3 - Q_1 = 66 - 41 = 25$$

C Percentiles and Percentile Rank

A measure that divided a ranked data set in increasing order into 100 equal parts is called percentile. The k^{th} percentile is denoted by P_k, where $k = 1,2,3,\cdots,99$. Therefore, the percentile k^{th} can be defined as a value in a ranked data set such that about $k\%$ of the measurements are smaller than the value P_k, and about $(100 - k)\%$ of the measurements are greater than the value of P_k.

Remark

Percentiles are not very useful for small data set.

Definition 3.3.4

The approximate value of the k^{th} percentile, denoted by P_k is

$$P_k = \text{value of the } \left(\frac{k \cdot n}{100}\right)^{th} \text{ term in a ranked data set}$$

Where k denotes the number of the percentile and n represents the sample size.

Note that, when we find the k^{th} percentile, we must first rank the data set and find the value $\frac{k \cdot n}{100}$. If the computed value is not a whole number, then round it to the next whole number

Let us consider the following examples

Example 5

Find the value corresponding to the 25^{th} percentile in the given data set:

18, 15, 12, 6, 8, 2, 3, 5, 20, 10

Solution: First, we rank the given data in increasing order.

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

Compute the value

$$\frac{k \cdot n}{100} = \frac{(25)(10)}{100} = 2.5$$

Since the value 2.5 is not a whole number, then round it to the next whole number, so the third number is P_{25}, and hence we reach $P_{25} = 5$.

Example 6

Find the value corresponding to the 60^{th} percentile in the given data set:

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

Solution: Since the data given already ranked, then we compute

$$\frac{k \cdot n}{100} = \frac{(60)(10)}{100} = 6$$

Since the value is whole number, in this case, use the average value between 6^{th} and 7^{th} values, that is

$$P_{60} = \frac{10 + 12}{2} = 11$$

Sometimes when we are interested in computing the percentage of values in the data set that are less than a specified value, a tool that used to make this is called **percentile rank**. Therefore the percentile rank of x_i gives the percentage of values in the data set that are less than x_i.

Definition 3.3.5

The percentile rank of a given x-valu can be computed by the formula

$$\text{Percentile rank of } x_i = \frac{\text{Number of values less than } x_i + 0.5}{\text{Total number of values in the data set}} \cdot 100\%$$

Let us consider an example

Example 7

Find the percentile rank for the score 8 in the given data set:

5, 6, 2, 3, 8, 12, 10, 15, 18, 20

Solution: First, we rank the given data in increasing order

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

Number of values less than 8 is equal to 4 and the total number of values is 10, so the percentile rank of 8 is

$$\frac{4 + 0.5}{10} \cdot 100\% = 45\%$$

That is, the score 8 is better than 45%of the class.

Exercises 3.3

1 A final examination for statistics course has a mean of 85 and a standard deviation of 5. Find the corresponding standard score for each raw score

a 88

b 65

c 79

d 90

e 93

2 The average teacher's salary in a university is SAR 10,000. If the standard deviation is SAR 5,000, find the salaries corresponding to the following z-scores

a 2

b 2.5

c -1

d -2

e 0

3 Which has a better relative position: a score of 70 on a statistics test with a mean of 60 and a standard deviation of 10; or a score of 35 on an accounting test with a mean of 30 and a variance of 16.

4 Which score indicates the highest relative position

a A score of 3.2 on a test with $\overline{x} = 4.6$ and $s = 1.5$

b A score of 3.2 on a test with $\overline{x} = 800$ and $s = 200$

c A score of 3.2 on a test with $\overline{x} = 50$ and $s = 5$

5 If you have the following distribution. By constructing a percentile graph for the following distribution

Class	(f_m)
10 – 19	5
20 – 29	15
30 – 39	20
40 – 49	40
50 – 59	10
Sum	**90**

a 30^{th} percentile
b 50^{th} percentile
c 75^{th} percentile
d 90^{th} percentile

6 Find the approximately ranks of these scores.

a 2 b 2.5 c -1 d -2 e 0

By constructing a percentile graph for the following distribution

Class Boundaries	(f_m)
196.5 – 217.5	5
217.5 – 238.5	17
238.5 – 259.5	22
295.5 – 280.5	48
280.5 – 301.5	22
301.5 – 322.5	6
Sum	**120**

7 Find the percentile rank for each value in the following data

1.1, 1.7, 1.9, 2.1, 2.2, 2.5, 3.3, 6.2, 6.8, 20.3

8 Find the z-score for each test, and state which is higher

Test	x	$\bar{x}$	s
I	58	40	9
II	70	60	8

9 The weekly salaries (in dollars) of sixteen government workers are listed below. Find the third quartile (Q_3).

492	794	545	833
506	747	611	798
690	876	450	589
709	473	936	527

10 The test scores of 19 students are listed below. Find the inter quartile range.

91	47	86	68	59	63	97
55	90	79	82	83	53	88
75	42	92	94	66		

11 The weekly salaries (in dollars) of sixteen government workers are listed below. Find the first quartile (Q_1).

690	592	813	660	728	559	473	600
517	665	685	458	538	787	500	826

12 Find the percentile rank for the data value 55 in the data set

55 38 30 66 67 68 44

13 Find the percentile rank for the data value 14 in the data set

4	6	14	10	4
10	18	18	22	6
6	18	12	2	18

14 The weights (in pounds) of 30 newborn babies are listed below. Find P_{16}

5.5	5.7	5.8	5.9	6.1
6.1	6.4	6.4	6.5	6.6
6.7	6.7	6.7	6.9	7.0
7.0	7.0	7.1	7.2	7.2
7.4	7.5	7.7	7.7	7.8
8.0	8.1	8.1	8.3	8.7

15 The scores of 32 students are listed below. Find P_{46}

32	37	41	44	46
48	53	55	56	57
59	63	65	66	68
69	70	71	74	74
75	77	78	79	80
82	83	86	89	92
95	99			

16 Any of the following values (A, B, C or D) corresponding to the 25th percentile in the given data set ?

1, 5, 2, 6, 8, 2, 3, 5, 10, 13

A) 5 B) 2 C) 3 D)13

17 Any of the following values (A, B, C or D) is the interquartile range for the following data?

1, 1, 3, 5, 6, 9, 7

A) 3 B) 1 C) 7 D) 6

18 Any of the following values (A, B, C or D) is the P_{60} in the following given data set ?

1, 3, 2, 4, 6, 10, 13, 15, 12, 14

A) 6 B) 11 C) 10 D) 14

19 Any of the following values (A, B, C or D) is the percentile rank for the score 10 in the following given data set ?

3, 4, 2, 3, 8, 13, 10, 25, 1, 30

A) 65% B) 25% C) 10% D) 30%

20 A final examination for statistics course has a mean of 15 and variance 4. Then, any of the following values (A, B, C or D) correspond to standard score for the score 5 ?

A) 5 B) 2.5 C) -2.5 D) -5

3.4 Box-and-Whisker Plot

The graph presentation of data using the five measures: the median (Q_2), Q_1 and Q_3 , the smallest and the largest values in a data set, is called a **box plot**. A box plot is a tool that used to illustrate some features in the data set such as the spread, the outlier, and the skewness of a data set.

Definition 3.4.1

A box plot is a graph of a data set obtained by drawing a horizontal line from the minimum data value to Q_1, drawing a horizontal line from Q_3 to the maximum data value, and drawing a box whose vertical sides pass through Q_1 and Q_3 with a vertical line inside the box passing through the median Q_2

Moreover the box-and-whisker plot can help us to detect outliers in a data set.

Definition 3.4.2

An outlier is an extremely high or an extremely low data value, when compared with the data values.

Note that, we can identify an outlier if it is greater than the highest fence:

$$Q_3 + (1.5)\, IQR$$

or less than the lowest fence:

$$Q_1 - (1.5)\, IQR$$

Let use consider an example

Example 1

Construct a box-and-whisker plot and find any outlier if there exists for the following data.

75, 69, 84, 137, 74, 104, 81, 90, 94, 144, 79, 98.

Solution:

Step1: First, we rank the given data in increasing order

$$\underbrace{69\ 74\ 75\ 79\ 81\ 84}_{\text{values less than the median}}\ \underbrace{90\ 94\ 98\ 104\ 137\ 144}_{\text{values greater than the median}}$$

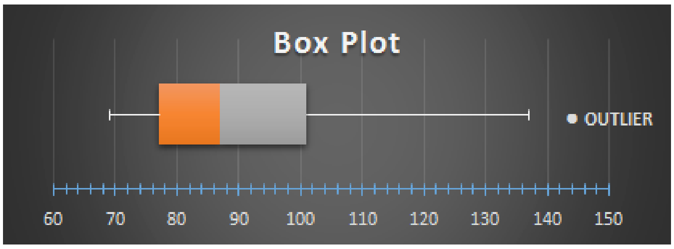
Step2: The three quartiles are:

$$Q_2 = \frac{84+90}{2} = 87,\ Q_1 = \frac{75+79}{2} = 77,\ Q_3 = \frac{98+104}{2} = 101$$

To find the outliers, we compute the following
$IQR = Q_3 - Q_1 = 101 - 77 = 24,\ 1.5(IQR) = (1.5)(24) = 36,$
$Q_1 - 36 = 77 - 36 = 41$ lowest fence, $Q_3 + 36 = 101 + 36 = 137$ highest fence.
Thus, the values less than 41 or greater than 137 are the outliers. Here the only outlier is 144.

Step3: Smallest value within the two inner fences is 69, and largest value within the two inner fences is 137.

Step4: Draw a vertical line and make the levels on it such that all values in the given data. To the right of the horizontal line, draw a box its bottom at the position of the first quartile and the top side at the position of the third quartile. Inside the box, draw a vertical line at the position of the median. Therefore, we have:



Exercises 3.4

1 Check each data set for outliers

a 16, 18, 22, 19, 3, 21, 17, 20

b 24, 30, 45, 31, 16, 18, 19, 15, 20, 17

c 14, 27, 18, 13, 16, 36, 19, 21, 15

d 61, 84, 82, 97, 88, 68, 100, 85

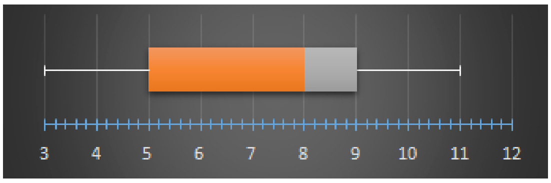
e 122, 118, 145, 119, 125, 116

2 The following data give the number of new cars sold at a dealership during a 20-day period. Make a box-and-whisker plot.

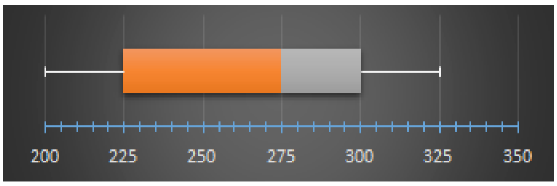
8	5	12	3	9
10	6	12	8	8
4	16	10	11	7
7	3	5	9	11

For Exercises 3 through 6, use each boxplot to identify the maximum value, minimum value, median, first quartile, third quartile, and interquartile range.

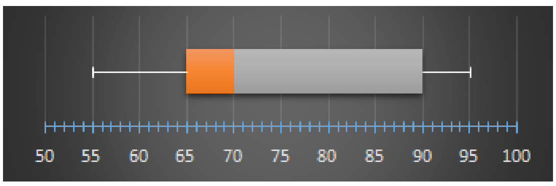
3



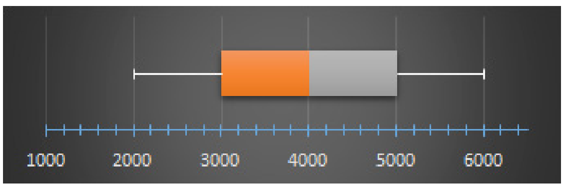
4



5



6



7 Construct a boxplot for the data below and identify the outliers

1.7	2.0	2.2	2.2	2.4	2.5	2.5	2.5
2.6	2.6	2.6	2.7	2.7	2.7	2.8	2.8
2.8	2.9	2.9	2.9	3.0	3.0	3.1	3.1
3.3	3.6	4.2					

8 Identify the outliers, if any, for the given data

35	41	56	65	67	68	70	73
75	77	78	82	87	90	99	

9 The weekly salaries (in dollars) of sixteen government workers are listed below

690	586	813	646	728	554	465	621
491	677	685	360	524	787	476	986

Find the outliers for the data above.

10 The weights (in pounds) of 18 randomly selected adults are given below

131	142	186	156	178	120	127	112
174	162	167	165	132	235	92	161
199	150						

Find the outliers for these data.

11 Construct a box-and-whisker plot and find any outlier if there exists for the following data.

75, 25, 84, 112, 74, 104, 81, 90, 94, 120,
79, 98.

SPSS Statistical Applications

Central tendency, variation and position measurements

This section of statistical applications discusses how students can compute values for central tendency measurements, variation measurements and position measurements using SPSS in order to make descriptive statistics about data, data could be either for a sample or a population. Given data could be ungrouped or grouped in defined classes. Moreover, students will learn how to plot a box-plot as an application of examples illustrated in section 3.4.

Example: Suppose we have the following data that represents a sample of (30) smoker males and the values are the ages of them:

45	37	53	32	61	27	39	44	50	58
54	73	35	23	16	72	40	51	25	25
25	37	42	43	28	26	24	65	60	49

Need, to find:

a) Mean, median, mode.

b) Minimum value, Maximum value, range, standard deviation, variance and the first and third quartiles denoted: Q1 (or P25), Q3 (or P75).

c) Percentiles 10th, 66th and 92th

d) From the results, compute the midrange (MR), Coefficient of variation (CV) and the Inter-quartile range (IQR).

e) From the results, compute Z-score of age 44.

f) Plot a boxplot for the given data and identify outlier(s) if exist(s).

Solution: To do this; first create a new data SPSS file, save it with name "data3.1" for example, and using the variable view save the variable with a name say "Age" which is scale and numeric with 0 decimals, after doing this we get the file as in figure 1 .

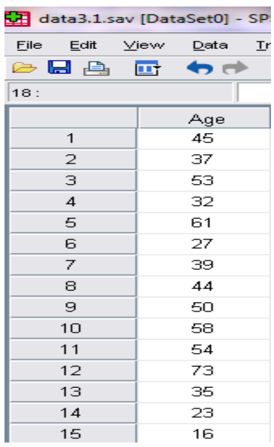


Figure1. Snapshot of the data file created

Now, to solve part (a) and part (b), follow the steps:

Analyze →Descriptive statistics →Frequencies →select variable “age” →press on the arrow to move it into variables box → press statistics button→ Select the central tendency and variation measurement needed and choose

The above command can be viewed as follows:

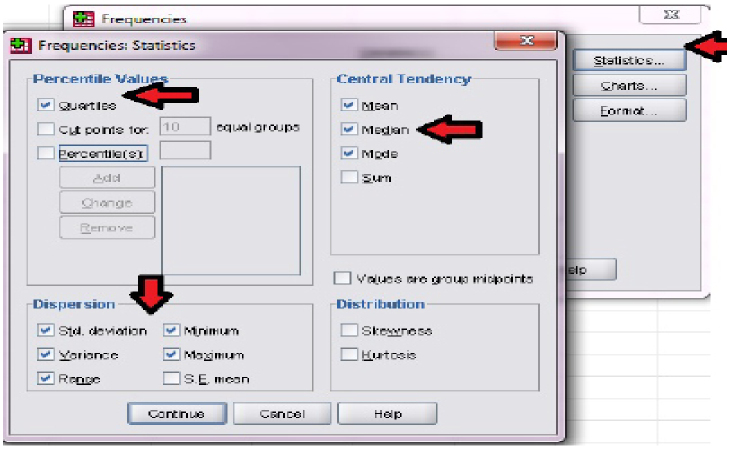


Figure2. Measurements of central tendency, variation and position

The results in figure 3:

The resulted plot will be as follows:

Frequencies

```
[DataSet1] I:\Statistics book\SPSS
           files\Ch#3\data3.1.sav
```

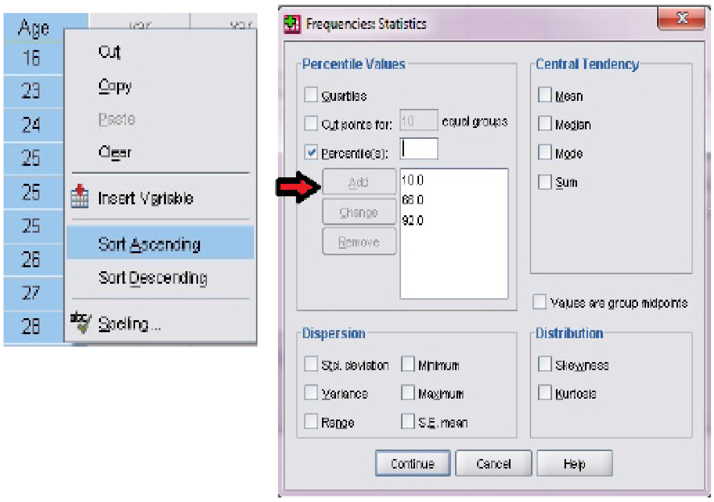
Statistics

Age of smokers

N	Valid	30
	Missing	0
Mean		41.97
Median		41.00
Mode		25
Std. Deviation		15.448
Variance		238.654
Range		57
Minimum		16
Maximum		73
Percentiles	25	26.75
	50	41.00
	75	53.25

Figure3. Measurements calculation

c) To compute the 10th, 66th and 92th, one can use the same command as in previous parts, but first you need to sort the data ascending as explained before, and check percentiles, then input the values (10, 66, 92) separately, after each input press add, then continue and finally ok. We get;



Frequencies

[DataSet2] I:\Statistics book\SPSS files\Ch#3\data3.1.sav

Statistics

Age of smokers

N	Valid	30
	Missing	0
Percentiles	10	24.10
	66	49.46
	92	68.64

Figure4. Percentiles calculation

As shown in figure 4, P10 = 24.10, P66= 49.46 and P92=68.64.

d) To compute the midrange (MR), Coefficient of variation (CV) andinterquartile range (IQR). Using results in figure 3: using equation in section 3.1, then

$$MR = \frac{Lowest + highest}{2} = \frac{16+73}{2} = 44.5$$

Using equation in section 3.2, then

$$CV = \frac{s}{x}.100\% = \frac{15.448}{41.97}.100\% = 36.80\%$$

Using equation in section 3.3, then

$$IQR = Q_3 - Q_1 = 53.25 - 26.75 = 26.50$$

e) To compute the z-score for age =44, one can use the result from parts (a) and (b) and using the equation illustrated in section 3.3:

$$z = \frac{x - \text{mean}}{\text{standard deviation}}$$

With x=44, mean=41.97, and standard deviation =15.448, thus

$$z = \frac{44 - 41.97}{15.448}$$

f) to graph a boxplot using SPSS, the following steps should be conducted;

> Graph→Legacy dialogs→boxplot→choose"simple"→summaries of separate variables → define→ in box represents field select the "age" variable→ OK

Figures 5.a and 5.b explains the previous command:

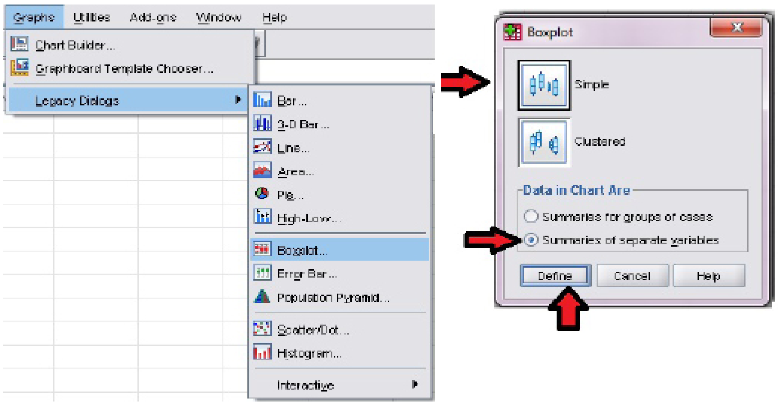


Figure5.aSteps of plotting Boxplot of variable "age"

Figure5.aSteps of plotting Boxplot of variable "age"

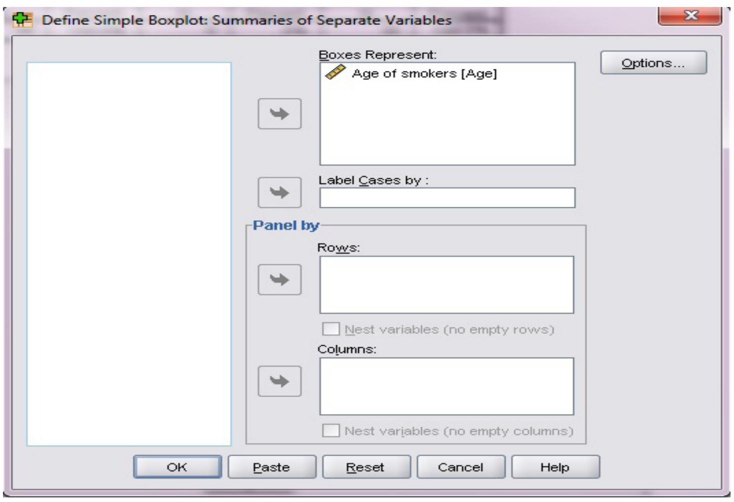


Figure5.aSteps of plotting Boxplot of variable "age"

And we get;

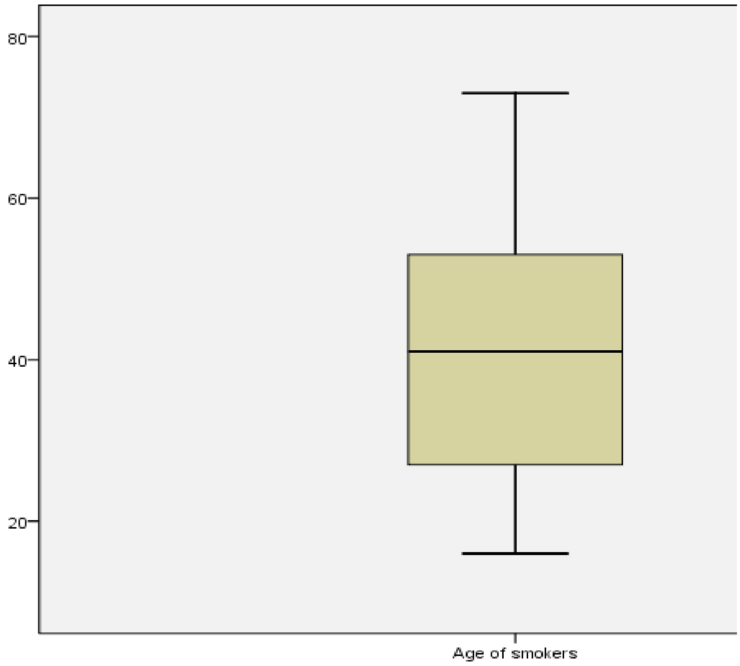


Figure 6. The resulted box-plot of variable "age"

As shown in figure 6, there is no outlier value in the given data. And at the end of this section we learned how to mix between information illustrated in lectures like equations of measurements and using the results in statistical environment of SPSS in order to solve statistical issues face us.

CHAPTER 4

Basic Concepts of Probability and Counting Rules

OBJECTIVES

1. Realize basic concepts about random experiments.
2. Obtain the space of elementary events of a random experiment.
3. Find the probability of events using the addition rules.
4. Find the probability of compound events using the multiplication rule.
5. Find the conditional probabilities of an event.
6. Applying the Bayes theorem to find probability of an event.
7. Find the probability of an event using counting rules.

Recall the definition of Statistics science as explained in chapter 1, that it's a part of science that concerning about the collecting and analyzing of raw data to get considerable results that may be used in the process of making decisions, one of the statistical techniques used in order to perform studies is an experiment; either observational or random experiment, differences between these two types and when to use each type are also illustrated in section 1.4. Mainly, usually resort to the use of random experiments when the researcher is interested in knowing the chance of occurrences for specified results. Section 4.1 discusses the process of performing a random experiment and the basic concepts used in such experiments. Before we go in depth, assume the following example. A student may ask an instructor about the chances of either (he /she) passing a course, getting at least B degree in the midterm exam or did not do well in the midterm.

In this chapter we want to discuss the basic concepts of probability that covering the random experiments, the space of elementary events, probability axioms, and rules of probabilities. To understand the rules of probabilities we need to know about Venn's diagrams to illustrate the relationship between events A, B, C, ... and S for a random experiment. Moreover, we will learn the rule for counting to compute the probability of any event in any space of elementary events.

4.1 Experiment, Outcome, and Space of Elementary Events

In a company that manufacture mobile phones, the quality control inspector of phones at the company may capture a phone from the production line to check whether it is good or defective. Inspector of phone is an example of a random experiment. The result of his inspection will be that the captured phone is either "good" or "defective." Each of these two observations is called an outcome of the random experiment, and these outcomes together represent the set of outcomes for such experiment.

Depending on the results of an experiment, there are two types of experiments, the regular experiments and the random experiments. To distinguish between these types consider an example, in chemistry science the results of any interaction are fixed (no change of results) under constant circumstances,. While in random experiments the results are not fixed; they may change in each trial despite the consistency of the surrounding circumstances. Thus, there is a factor

of chance, in other words, probability is a tool to guess the chance of an event will be happen.
The basic concepts of probability are defined as in the following

Definition 4.1.1

A random experiment is the process by which an observation (or measurement) is obtained , but say for sure without knowing the outcome in advance.

Definition 4.1.2

An elementary event is an outcome of a single trial of a random experiment.

Definition 4.1.3

A space of elementary events is the set of all elementary events (possible outcomes) of a random experiment, denotes by S.

For the example of inspector phones the space of elementary events can be written as follows:

$$S = \{\text{Good, Defective}\}$$

Other examples on random experiments, elementary events (outcomes), and their space of elementary events are list in the following table.

Experiment	Outcomes	Sample space
Toss one fair coin	Head, Tail	$S = \{\text{Head}, \text{Tail}\}$
Roll a fair die once	1, 2, 3, 4, 5, 6	$S = \{1,\ 2,\ 3,\ 4, 5, 6\}$
Toss two fair coins	$HH,\ HT,\ TH,\ TT$	$S = \{HH,\ HT,\ TH,\ TT\}$
Answer a True/ False question	True, False	$S = \{\text{True}, \text{False}\}$

One conclusion from the above table is that, the space of elementary events involves all outcomes of a random experiment. An interesting thing in statistics is that, the data could be represented using graphical diagrams to understand the statistical concepts, in a random experiment two common diagrams are used in order to draw the space of elementary events, namely; Venn diagram and the tree diagram. A Venn's diagram uses a geometric shape like a rectangle or a circle that contains all the possible outcomes for a random experiment. While in a tree diagram method branches are used to represent the outcomes of a random experiment.

Example 1

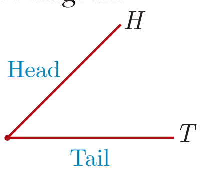
A coin is tossing once. Find the sample space, and draw the Venn and tree diagrams for this experiment.

Solution: This experiment has two possible outcomes: head (H) and tail (T), the space of elementary events is given by

$$S = \{H, T\}$$

To draw a Venn's diagram, we draw a rectangle as shown in figure 4.1(a), and mark two points inside this rectangle that represent the two outcomes, H and T. The rectangle is labeled S because it represents the space of elementary events. Also, Figure 4.1(b) shows the representation of the outcomes using the tree diagram

4.1(a)



4.1(b)

Example 2

Find the space of elementary events and draw a tree diagram for rolling one die.

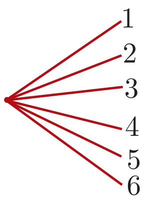
Solution:

This experiment has six possible outcomes:

1, 2, 3, 4, 5, and 6.

Therefore the space of elementary events is given by

$$S = \{1, 2, 3, 4,5,6\}$$



Example 3

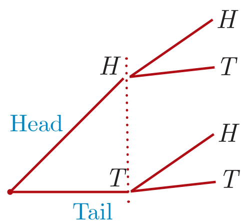
Find the space of elementary events and draw a tree diagram for tossing two different coins at once.

Solution: This experiment has four possible outcomes (four elementary events), and we can obtain these outcomes as follows:

The first possible outcome is (HH) this mean that we get a head (H) on the first coin and a head (H) on the second.

The second possible outcome is (HT) this mean that we get a head (H) on the first coin and a Tail (T) on the second.

The third possible outcome is (TH) this mean that we get a tail (T) on the first coin and a head (H) on the second.



The fourth possible outcome is (TT) this mean that we get a tail (T) on the first coin and a tail (T) on the second. Thus the space of elementary events is given by

$$S = \{HH,\ HT,\ TH,\ TT\}$$

Example 4

Find the space of elementary events for tossing one coin and rolling a die together one time.

Solution: This experiment has twelve possible outcomes, and we can obtain these outcomes as follows:

The first six possible outcomes are the pairs (H, 1), (H, 2), (H, 3), (H, 4), (H, 5), and (H, 6) this means that we get a head (H) on the coin corresponds with the possibility of get one aspect of the die 1, 2, 3, 4, 5, and 6. The second six possible outcomes are the pairs (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), and (T, 6) this means that we get a head (T) on the coin corresponds with the possibility of get one aspect of the die 1, 2,3,4,5, and 6. Thus the space of elementary events is

$$S = \left\{ \begin{matrix} (H,1),(H,2),\ (H,3),\ (H,4),\ (H,5),\ (H,6) \\ (T,1),\ (T,2),\ (T,3),\ (T,4),\ (T,5),\ (T,6) \end{matrix} \right\}$$

Example 5

Find the space of elementary events for tossing a coin three times.

Solution: This experiment has eight possible outcomes, and we can obtain these outcomes as follows:

The first possible outcome is the triple (H, H, H) this means that we get a head (H) in the three times. The second possible outcome is the triple (H, H, T) this means that we get a head (H) in the first and second times and get a tail in the third time.

Continue with same manner we can construct the space of elementary events to be

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Previous examples discuss the representation of outcomes using the Tree diagrams, in some cases a subgroup of outcomes may gathered in a subset of the space of elementary events and the outcomes inside this subset share a common characteristic, this subset is called an event, more details about events are in the next subsection.

Concept of Event

Consider an experiment of rolling a die once, the space of elementary events would be the numbers 1,2,3,4, 5 and 6, assume that O is an event that consists of odd numbers outcomes in the performed experiment, then O contains the numbers 1,3 and 5, Moreover, if A is an event assumed to contain the resulted outcome equal to 7, then A is an empty event, since it's impossible to get such outcome in our experiment, when an event consists of no outcomes and represents an empty subset of the sample space, in other words, it's impossible to get outcomes under certain conditions, such event is named by "phi" in the references and denoted by ϕ .

Definition 4.1.4

An event is a subset of the space of elementary events of a random experiment and denoted by capital letter $(A, B,..)$

Definition 4.1.5

A simple event is an event which consists of only a single elementary event of the space of elementary events (single outcome of the random experiment).

Referring to example 5 for tossing a coin three times, each of the eighth outcomes (elementary events) HHH, HHT, HTH, HTT, THH, THT, THH and TTT for this experiment generates a simple event. These eighth events can be denoted by $E_1, E_2, E_3, \ldots,$ and E_8 respectively. Thus

$$E_1 = \{HHH\}, E_2 = \{HHT\}, E_3 = \{HTH\} E_4 = \{HTT\},$$
$$E_5 = \{THH\}, E_6 = \{THT\}, E_7 = \{THH\}, E_8 = \{TTT\}.$$

A Compound Events

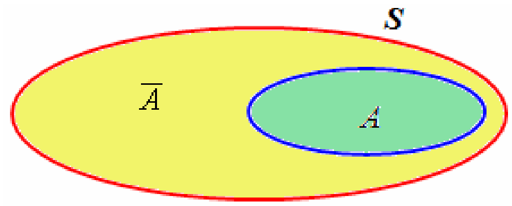
One encounters in many cases events are result of algebraic (or logic) operations between events. These events are called compound events.

Definition 4.1.6

A compound event is an event which generated by algebraic operations between events from the space of elementary events.

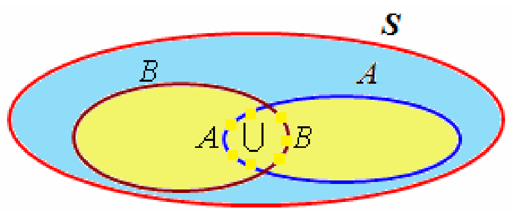
Example 6

We conceder S as a space of elementary events, and let A be an event from S. Then $S \backslash A$ is an event occurs when the elementary event (the result of the random experiment) does not belong to the event A, and called complementary event of A and denoted by $\overline{A}$, e.g. $\overline{A} = S \backslash A$. The following Figure gives us the Venn's diagram of $\overline{A}$



Example 7

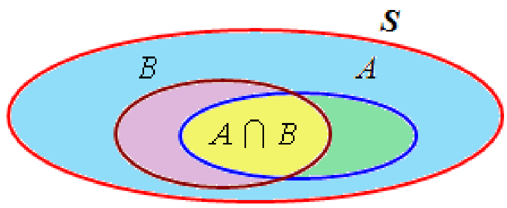
We consider S as a space of elementary events, and let A and B be events from S. Then $A \cup B$ is an event occurs when the result of random experiment (elementary event) belong to the event A or B or to the both events. The following Figure gives us the Venn's diagram of $A \cup B$.



And for $B = \overline{A}$ one becomes $A \cup \overline{A} = S$, therefore S is called the certain event.

Example 8

We consider S as a space of elementary events, and let A and B be events from S. Then $A \cap B$ is an event occurs when the result of random experiment (elementary event) belong to the event A and B together. The following Figure gives us the Venn'^s diagram of $A \cap B$.



And for $B = \overline{A}$ one becomes $A \cap \overline{A} = \phi$, therefore ϕ is called the impossible event.

Example 9

A die tossed two times, find the following events.

a. The sum of any two outcomes is 3.

b. The sum any two outcomes is lees than 4.

c. The sum any two outcomes is more than 10.

Solution:

$$(a)\ E_1 = \{(1,2),(2,1)\},\ \ (b)\ E_2 = \{(1,1),(1,2),(2,1)\}$$

$$(c)\ E_3 = \{(5,6),(6,5),(6,6)\}$$

Note that

a) From the last example we note that each event can be seen as a composed event consists of simple events. Because we can write:

$$E_1 = \{(1,2)\} \bigcup \{(2,1)\}$$
$$E_2 = \{(1,1)\} \bigcup \{(1,2)\} \bigcup \{(2,1)\}$$
$$E_3 = \{(5,6)\} \bigcup \{(6,5)\} \bigcup \{(6,6)\}$$

b) If we have $A \cap B = \phi$ for two events A and B from S. Then we can say A and B are mutually exclusive events.
c) Let $A_1, A_2, ..., A_n$ be events from S. Now if we have $A_i \bigcap A_j = \phi$ for $i \neq j$ and $i, j = 1,2,......, n$, then one say $A_1, A_2, ..., A_n$ are pairwise mutually exclusive.

In experiments repeated many times, one may ask about the probability of an event consists of outcomes that their sum equals to 5, section 4.2 discusses the methodology of calculation the probability of events.

Exercises 4.1

1 What is the difference between an elementary event and a compound event?

2 Find the space of elementary events and draw a tree diagrams for rolling two dice

3 Find the space of elementary events and draw a tree diagrams for tossing one coin and die.

4 Find the space of elementary events and draw a tree diagrams for tossing a coin four times.

5 A die tossed two times, find the following compound events.

a The sum of any two outcomes is 5

b The sum of any two outcomes is less than 2.

c The sum of any two outcomes is more than 7.

6 Tow different dice tossed one time, then find the following compound events.

a The sum of any two outcomes is odd.

b The sum of any two outcomes is even.

c The sum of any two outcomes is prime.

d The sum of any two outcomes is less than 5.

e The sum of any two outcomes is more than 5.

7 Find the space of elementary events for tossing two coins and rolling a die together.

8 A die tossed two times, find the following compound events.

a The sum of any two outcomes is 2 or 1.

b The sum of any two outcomes is even and prime.

c The sum of any two outcomes is prime and odd.

d The sum of any two outcomes is less than 1.

e The sum of any two outcomes is less than or equal 1

9 A die tossed two times, find the following events and their number of outcomes (elementary events).

a The sum of any two outcomes is 1.

b The sum of any two outcomes is prime less than two.
c The sum of any two outcomes is prime and odd.
d The sum of any two outcomes is greater than 0.
e The sum of any two outcomes is less than or equal 1

10 A coin is tossing two times. Find the space of elementary events, and draw the tree diagram for this experiment.

11 A die tossed three times, find the following events and their number of outcomes (elementary events).

a The sum outcomes is 3.
b The sum of the first two components is prime and the third is odd.
c The sum of the first two components is prime and the third is even.
d The middle component is even.
e The third component is even.

12 A die tossed three times, find the following events and their number of outcomes (elementary events).

a The sum of the components is negative.
b The first two components is odd and the third is even.
c The sum of the first two components is odd and the third is prime less than 4.
d The sum of the first two components is odd and the third is prime less than or equal 4.
e The middle component is positive.

13 A die tossed two times, find the space of elementary events and following events.

a The sum of the components is zero.
b The first component is one and the second is even.
c The sum of the two components is 2.

14 A die tossed two times, then:

a If A is the event that the sum of the components equals to 8 what is the event $\overline{A}$?
b If B is the event that the first component equals to odd number, and C is the event that the second component leas than 4, what is the event $B \cup C$?

c If D is the event that the first component equals to odd number, and E is the event that the second component leas than 4, what is the event $D \cap E$?

15 A coin tossed four times, then:

a If A is the event that we get three heads what is the event $\overline{A}$?

b If A is the event that we get tails in the first and second times, and B is the event that we get heads in the first and second times, what is the event $A \cup B$?

c If A is the event that we get a tail in the last time, and B is the event that we get a head in the first time, what is the event $A \cap B$?

16 There are in a box two white, two black and four blue balls. We draw three balls in a consecutive with the repatriation, then:

a a If A is the event that we get three blue balls what is the event $\overline{A}$?

b b If A is the event that we get a black ball in the first time and a white ball in the first second time, and B is the event that we get a blue ball in the third time, what is the event $A \cup B$?

c If A is the event that we get a blue ball in the fist time, and B is the event that we get a black ball in the last time, what is the event $A \cap B$?

17 If you flip a coin three times, the possible outcomes are

A) $\{HHH\ HHT\ HTH\ HTT\ THH\ THT\ TTH\}$

B) $\{HH\ HT\ TH\}$

C) $\{HHH\ HHT\ HTH\ HTT\ THH\ THT\ TTH\ TTT\}$

D) $\{HHT\ HTH\ HTT\ THH\ THT\ TTH\ TTT\}$

18 A die tossed two times. If E is the event that represent the outcomes is 6, then any of the following events (A, B, C or D) is the event E?

A) $\{(2,2)\}$

B) $\{(1,5), (2,4)\}$

C) $\{(1,1),(2,2),(3,6)\}$

D) $\{(1,1)\}$

4.2 Calculating Probability

Probability function, which gives the likelihood of happening of an event, is denoted by P. For an event A, the probability that A will occur is a real number from the interval $[0,1]$ and denoted by $P(A)$, e.g. for S a space of elementary events we have:

$$P : \{A\ ;\ A \text{ event from S}\} \longrightarrow [0\ ,\ 1]$$
$$A \longmapsto P(A)$$

So we have for any event A from S:

$$0 \leq P(A) \leq 1$$

Definition 4.2.1

Probability of an event is a numerical value indicates to the likelihood that the event will occur.

The likelihood of occurrence for any event is restricted by some rules that should be satisfied, in order to be considered as a probability, here are the properties of probability.

Properties of Probability

The probability of an event either is never less than 0 or greater than 1 but it could equal to zero or one, and the sum of the probabilities of all simple events is always 1. In mathematical notation the probability function P verifying the following properties:

a) For the impossible event ϕ must be $P(\phi) = 0$.

b) If $A_1, A_2, \ldots\ldots, A_n$ pairwise mutually exclusive events from S, then must be:

$$P\left(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots\right) = P\left(A_1\right) + P\left(A_2\right) + \cdots + P\left(A_n\right) + \cdots$$

c) For the certain event S must be $P(S) = 1$.

From the last property we note that the sum of the probabilities of all simple events equal to 1.
Reconsider example 2 that rolled one fair die, the sample space and all simple events respectively; are

$$S = \{1,2,3,4,5,6\} \text{ and}$$
$$E_1 = \{1\},\ E_2 = \{2\},\ E_3 = \{3\}$$
$$E_4 = \{4\},\ E_5 = \{5\},\ E_6 = \{6\}$$

Notice that, the probability of each one of these simple events is $\frac{1}{6}$. The sum of all simple events in this example is closed to one, that is

$$\begin{aligned}\sum P(E_i) &= P(E_1) + P(E_2) + P(E_3) + P(E_4) + P(E_5) + P(E_6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= 1\end{aligned}$$

Conceptual Approaches to Probability

There are several conceptual approaches to calculate probabilities of events, and some of them:

a) The **classical definition of probability** (or the Laplace principle in probability)

b) The **relative frequency concept of probability**,

c) The **subjective probability concept**.

Here we will discuss the the first two approaches only.

A The Classical Definition of Probability

In a probability experiment and it is outcomes, we may have the same probability of occurrence outcomes. Such outcomes are called equally likely, and therefore we say that the elementary events have the same probability. Recall example 2, then all outcomes in this example are equally likely with chance equal to $\frac{1}{6}$, thus the elementary events have the same probability $\frac{1}{6}$.

The classical definition of probability is used to compute the probabilities of events for an experiment for which all outcomes are equally likely.

The probability of a simple event E and an event A can be calculated by applying the following formula respectively :

$$P(E) = \frac{1}{\text{Total number of outcomes}}$$

$$P(A) = \frac{\text{number of outcomes include in } A}{\text{Total number of outcomes}}$$

Using the notation, we write the classical definition of probability as follows

If A is an event of a space of elementary events S, then

$$P(A) = \frac{n(A)}{n(S)}$$

where $n(A)$ is the number of outcomes of A, and $n(S)$ is the number of outcomes of the space of elementary events S,

The following examples illustrate how the classical definition of probability can be applied

Example 1

Find the probability of obtaining a head and the probability of obtaining a tail for one toss of a coin.

Solution: The two outcomes, head and tail, are equally likely outcomes. If the events E_1 and E_2 represent the outcome are a head and a tail respectively, then

$$P(E_1) = \frac{1}{\text{Total number of outcomes}} = \frac{1}{2}$$

Similarly, we have

$$P(E_2) = \frac{1}{\text{Total number of outcomes}} = \frac{1}{2}$$

Example 2

Compute the probability of obtaining an odd number in one roll of a die.

Solution: The outcomes of this experiment are: 1, 2, 3, 4, 5, and 6. All these outcomes are equally likely. If A be an event that an odd number is observed, then the event A includes three outcomes: 1, 3, and 5; that is,

$$A = \{1,3,5\}$$

Note that $n(S) = 6$ and $n(A) = 3$. Hence

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

Example 3

If a family has three children, find the probability that two of the three children are boys.

Solution: The space of elementary events of this experiment is

$$S = \{BBB, BBG, BGB, GBB, GGB, GBG, BGG, GGG\}$$

If A is the event that include two boys of the three children, then A include three outcomes, that is

$$A = \{BBG, BGB, GBB\}$$

Note that $n(S) = 8$ and $n(A) = 3$. Hence

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{8}$$

In real life, the events of random experiments are not always equally likely, so that, it is needed to create another technique to compute probability in such experiments.

B Relative Frequency Concept of Probability

Some time the classical definition of probability is not suitable to apply to compute probabilities, this because the various outcomes for the corresponding experiments are not equally likely. For instance, assume that we are interest to compute the probability that the next car that comes out of an auto factory in a company is a "Black". The next car manufactured at an auto factory may be Black or may be not. The two outcomes, "it is a Black" and "it is not a Black," are not equally likely. If they were, then (approximately) half the cars manufactured by this company would be Black, and this is unacceptable.

To calculate probabilities, a data may be generated by performing the experiment a large number of times or we use past data. The relative frequency of an event is used as an approximation for the probability of that event. This method of assigning a probability to an event is called the **relative frequency concept of probability**.

The probability of an event A is equal it's relative frequency. That is

$$P(A) = \text{Therelative frequany} = \frac{\text{the frequany of } A}{\text{Total number of trials}} = \frac{f}{N}$$

Let us consider some examples

Example 4

For tossing a coin 100 times, if we get 20 heads and 80 tials. Find the probability of getting

a. A head

b. A tail

Solution: The two outcomes, head and tail, are equally likely. If the events H and T represent the outcomes are a head and a tail respectively, then

$$P(H) = \frac{20}{100} = 0.2 \text{ and } P(T) = \frac{80}{100} = 0.8$$

Note that according to example 4, the sum of the two probabilities equals to one, and both probabilities are greater than zero and less than one.

The probabilities calculated using relative frequencies may change almost each time an experiment is repeated. Therefore, we note that the method of relative frequency method does not suitable for calculating the probability of an event.

A third approach for calculating probabilities is the subjective probability concept and it is used in experiments that neither could be repeated as in the relative frequency nor could happen equally likely as in classical approach, for example, assume that a student who is taking a mathematics course at PY, will get B grade in this course.

Formula of Probability

We face many problems when finding the probability of more than one event. Consider the following two examples.

A. Choose a person randomly from one of the big markets, and we might wish to know, the probability that the person is a businessman or is a university student. In this case, there are three possibilities to consider:

i. The person is a businessman,

ii. The person is a university student,

iii. The person is both a businessman and a university student.

B. A person is selected randomly at the same markets there are university students, married, and unmarried. What is the probability that the person is a marred or an unmarried? In this case, there are only two possibilities:

a. The person is a married,

b. The person is not married.

The difference between the two examples A and B is that, in example A, the person selected can be a businessman and a university student at the same time. While in example B, the person selected cannot be both marred and not at the same time. In example B, the two events are said to be mutually exclusive; in example A, they are not mutually exclusive. Where we introduced this concept previously, and in order to provide clarification we illustrate the following example.

Example 5

Determine which events are mutually exclusive and which are not, when a single dice is rolled.

a. Getting an even number and getting an odd number at same time

b. Getting a 5 and getting an odd number at same time

c. Getting an even number and getting a number less than 6 at same time

d. Getting a number greater than 3 and getting a number less than 3 at same time

Solution:

a. The events are mutually exclusive, since the first event could be 2, 4, or 6 and the second event could be 1, 3, or 5.

b. The events are not mutually exclusive, since the first event is a 5 and the second can be 1,3, or 5. Hence, 5 exists in both events.

c. The events are not mutually exclusive, since the first event can be 2, 4, or 6 and the second can be 1, 2, 3, 4, or 5. Hence, 2 and 4 are contained in both events.

d. The events are mutually exclusive, since the first event can be 4, 5 or 6 and the second event can be 1, or 2.

The probability of two or more events can be determined by the well known addition rules. The first addition rule is used when the events are mutually exclusive.

Addition Rule 1

For any two mutually exclusive events A and B, the probability that A or B will occur is

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

To understand the addition rule 1 consider the following example 6

Example 6

In a sample of 100 people, 42 had type O blood, 44 had type A blood, 10 had type B blood, and 4 had type AB blood. Set up a frequency distribution and find the following probabilities for a person to have:

a. Type A or B blood.

b. Neither type A nor Type O blood.

Solution: The frequency distribution is illustrated in the following table

Type	Frequency
A	44
B	10
AB	4
O	42
Sum	**100**

a. Since having of blood O, A, B and AB are mutually disjoint events then we have

$$P(\mathrm{A} \cup \mathrm{B}) = P(\mathrm{A}) + P(\mathrm{B}) = \frac{44}{100} + \frac{10}{100} = \frac{54}{100} = \frac{27}{50}$$

b. $P(\text{Neither A nor O}) = P(\mathrm{B} \cup \mathrm{AB}) = P(\mathrm{B}) + P(\mathrm{AB})$

$$= \frac{10}{100} + \frac{4}{100} = \frac{14}{100} = \frac{7}{50}$$

When two events are not mutually exclusive, we must subtract one of the two probabilities of the outcomes that are common to both events, since they have been counted twice.

Addition Rule 2

If A and B are two events taken from the space of elementary events S and $A \cap B \neq \varnothing$, then

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

To understand the addition rule 1 consider the following examples

Example 7

A single card is drawn at random from an ordinary deck. Find the probability that it is either an ace or a black card.

Solution: Since there are 4 aces and 26 black cards, 2 of the aces are black cards namely the ace of spades and the ace of clubs. Therefore, let denotes to the event for aces and denotes to the event for black.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52} = \frac{7}{13}$$

Example 8

If A and B are two events selected from the sample space S where $P(A) = 0.6, P(B) = 0.7$ and $P(A \cap B) = 0.4$
Find $P(A \cup B)$

Solution: Applying the addition rule 2, we get

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 0.6 + 0.7 - 0.4 = 0.9$$

Complementary Events

We have previously introduced the concept of complementary event where we stated that two mutually exclusive events that taken together include all the outcomes for an experiment are called complementary events. Note that two complementary events are always mutually exclusive. Therefore we can write:

Result 1

For any event A, we have

$$P(\overline{A}) = 1 - P(A) \text{ and } P(A) = 1 - P(\overline{A})$$

where $\overline{A}$ is the complement of the event A.

Example 9

Find the probability of the following events

a. Probability that getting a 5 when rolling a die once.

b. Probability that not getting a 5 when rolling a die once.

c. Getting a weekend when selecting a day of the week.

d. Probability that not getting a weekend when selecting a day of the week

Solution:

a. The space of elementary events in this case is

$$S = \{1,2,3,4,5,6\}$$

If we denote to the event of getting a 5 by E (since it is a simple event), then $E=\{5\}$ and therefor

$$P(5) = \frac{1}{6}$$

b. If we denote to the event of getting a 5 by E, then the event of not getting a 5 is the complement of E that is $(\bar{E})$, so

$$P(\overline{E}) = 1 - \frac{1}{6} = \frac{5}{6}$$

c. The space of elementary events in this case is

S={Saturday,Sunday,Monday,Tuesday,Wednesday,Thursday, Friday}.

The weekend days are Saturday and Friday, if we denote to such days by A, then A={Saturday and Friday}, so

$$P(A) = \frac{2}{7}$$

d. If we denote to the event of getting a weekend day by A, the event of not getting a weekend day is the complement of A that is ($\overline{A}$), so using result 1, then the probability of not getting a weekend day is

$$P(\overline{A}) = 1 - P(A) = 1 - \frac{2}{7} = \frac{5}{7}$$

Exercises 4.2

1 A survey classified a large number of adults according to whether they were judged to need eyeglasses to correct their reading vision and whether they use eyeglasses when reading. The proportion falling into four categories shown in the table.

Judged to need eyeglasses	Used eyeglasses for reading	
	Yes	No
Yes	0.44	0.14
No	0.02	0.40

If a single adult is selected from the large group, find the probability of each event:

a The adult is judged to need eyeglasses.
b The adult needs eyeglasses for reading but does not use them.
c The adult uses eyeglasses for reading whether he or she needs them or not.

2 If a fair die is rolled one time, find the probability of getting

a A 5
b An odd number.
c A number less than 7.
d A number more than 8.
e A number greater than 3 or an odd number.

3 If two fair dice are rolled one time, find probability of getting the results

a A sum of 9.
b Doubles.
c A sum of 8 or 12.
d A sum is greater that 10.
e A sum is less than or equal to 5.

4 If one card is drawn from an ordinary deck, find the probability of getting these results

a A red.
b A club.
c A 5 or 8.
d A 3 or a spade.
e A black king.
f A red card or a 9.

5 A bowl contains three red and two yellow balls, two balls are randomly selected one after the other without replacement (returning) and colors recorded. Find the probability that

a The two balls are red.
b The two balls are yellow.
c One is red and the other is yellow.
d They have the same color.

6 A family has four children. Find the probability:

a All boys.
b All girls.
c All boys or girls
d Exactly two boys or two girls.
e At least one child of each gender.
f At most one child is a girl.

7 A breakdown of the sources of energy used in Saudi Arabia is shown below.

Oil:60% **Natural gas:** 35% **Others:** 5%

Choose one energy source at random. Find the probability that:

a Not oil.
b Natural gas or oil.

8 Human blood is grouped into four types. The percentages of Saudi with each type are listed below:

Type	Percentage
O	45%
A	40%
B	12%
AB	3%

Choose one Saudi at random. Find the probability that the person has type:

a O blood.
b Type A or B.
c Does not have type O or A.

9 Of the top 10 cars based on gas mileage, 4 are Hondas, 3 are Toyotas, and 3 are Volkswagens. Choose one car at random. Find the probability that the car is

a Japanese.
b Japanese or German.

10 At a used-book sale, 200 books are adult books and 320 are children's books.Of adult books, 60 are fiction while 200 of the children's books are fictions. If a book is selected at random. Find the probability that it is

a Nonfiction.

b An adult book or children's nonfiction book.

11 The frequency distribution shown here illustrates the number of medical tests conducted on 60 randomly selected emergency patients.

Number of tests performed	Number of patients
0	24
1	16
2	4
3	6
4 or more	10

If a patient is selected at random, find the probabilities:

a The patient has had exactly 2 tests done.

b The patient has had at least 2 tests done.

c The patient has had at most 2 tests done.

d The patient has had 3 or fewer tests done.

e The patient has had 1 or 2 tests done.

12 Three cable channels (6, 8 and 10) have quiz shows, comedies, and dramas. The number of each is shown here

Types of show	Channel 6	Channel 8	Channel 10
Quiz show	10	4	2
Comedy	5	4	16
Drama	8	7	4

If a show is selected at random, find the probabilities:

a The show is a drama or a comedy.

b The show is shown on channel 10 or a quiz show.

13 In a statistics class there are 18 juniors and 10 seniors, 6 of the seniors are female, and 12 of the juniors are males. If a student is selected at random, find the probability of selecting the following

a A junior or a female.

b A senior or a male.

c A junior or a senior.

14 A recent study of 200 nurses found that of 125 female nurses, 56 had bachelor's degree; and of 75 male nurses, 34 had bachelor's degree. If a nurse is selected at random, find the probability that the nurse is

a A female nurse with a bachelor's degree.

b A male nurse.

c A male nurse with a bachelor's degree.

15 At a convention there are 7 mathematics instructors, 5 computer science instructors, 3 statistics instructors, and 4 science instructors. If an instructor is selected. Find the probability of getting a science or a math instructor.

16 The probability of obtaining a head for one toss of a coin is

A) 0.5 B) 0.25 C) 0.75 D) 1

17 If A and B are two events selected from the space of elementary events S where $P(A)=0.2$, $P(B)=0.3$ and $P(A \cap B) = 0.4$ Find $P(A \cup B)$

A) 0.3 B) 0.2 C) 0.05 D) 0.1

18 If $P(A)=0.03$, then $P(\overline{A})$

A) 0.97 B) 0.79 C) 0.01 D) 0.2

19 If $S=\{E_1.E_2,E_3\}$ and assume that $P(E_1)= P(E_3) =0.3$, then any of the following values (A, B, C or D) equal to $P(E_2)$?

A) 0.04 B) 0.4 C) 0.41 D) 0.3

4.3 Multiplication Rules and Conditional Probability

This section introduces the multiplication rules and conditional probability.

A The Multiplication Rules

The multiplication rules can be used to find the probability of two or more events that occur in sequence. For example, if you toss a coin and then roll a die, you can find the probability of getting a tail on the coin and on 3 on the die. These two events are said to be independent since the outcome of the first event (tossing a coin) does not affect the probability outcome of the second event (rolling a die).

Definition 4.3.1

Two events A and B are independent events if the fact that A occurs does not affect the probability of B occurring.

Let us consider example

Example 1

If you toss a coin and then roll a die, you can find that the probability of getting a head on the coin and 4 on the dice. Discuss the independency of the events.

Solution: These two events are said to be independent since the outcome of the first event getting a head in tossing a coin does not affect the probability of the second event getting 4 on rolling a die.

To find the probability of two independent events that occurred in sequence, we can use the following multiplication rule

Multiplication Rule 1

When two events are independent, the probability of both occurring is

$$P(A \cap B) = P(A) \cdot P(B)$$

Note that to find the probability of two independent events that occurred in sequence, we must find the probability of each event separately and then multiply the answers.

Let us consider an example

Example 2

A fair coin is flipped and a fair die is rolled. Find the space of elementary events S and then calculate the probability of getting a head on the coin and 3 on the die.

Solution: The space of elementary events of this random experiment is given by:

$$S = \begin{Bmatrix} (H,1),(H,2), & (H,3),(H,4), & (H,5),(H,6), \\ (T,1),(T,2), & (T,3), & (T,4),(T,5),(T,6) \end{Bmatrix}$$

Now we can calculate the probability of getting a head on the coin and 3 on the die as follows:

Method 1: Let A the event that is a head on a coin, and B the event that is 3 on the die, and then we have:

$$A = \{(H,1),(H,2),(H,3),(H,4),(H,5),(H,6)\}$$

So we have:

$$\begin{aligned} P(A) &= P(\{(H,1),(H,2),(H,3),(H,4),(H,5),(H,6)\}) \\ &= \frac{6}{12} = \frac{1}{2} \end{aligned}$$

And

$$B = \{(H,3),(T,3)\}$$

So we have:

$$\begin{aligned} P(B) &= P(\{(H,3),(T,3)\}) \\ &= \frac{2}{12} = \frac{1}{6} \end{aligned}$$

Therefore the event that a head on the coin and 3 on the die is:

$$A \cap B = \{(H,3)\}$$

So we have the probability of getting a head on the coin and 3 on the die equal to:

$$P(A \cap B) = P(\{(H,3)\}) = \frac{1}{12}$$

We note that A and B are independent because:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$

Method 2: Let C the event that is a head on a coin and 3 on the die, then we have:

$$C = \{(H,3)\}$$

So we have:

$$P(C) = P(\{(H,3)\}) = \frac{1}{12}$$

Example 3

An urn contains 5 red balls, 4 blue balls and 3 white balls. A ball is selected at random and its color noted, then it returned to the urn (replaced). A second ball is selected at random and its color noted. Find the probability of each of these events

a. Selecting 2 red balls.

b. Selecting one red ball and then one white ball.

Solution: :

For (a): Let A and B be an event that we get a red ball in the first and second selection respectively, and then we have:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{5}{12} \cdot \frac{5}{12} = \frac{25}{144}$$

For (b): Let C be an event that we get a white ball in the second selection, and then we have:

$$P(A \cap C) = P(A) \cdot P(C) = \frac{5}{12} \cdot \frac{3}{12} = \frac{15}{144}$$

This mean that the two events A and B are independent.

Example 4

If a fair die is rolled twice. Find the space of elementary events S and then calculate the probability of getting 3 on the first and 5 on the second.

Solution: : The space of elementary events of this experiment is given by:

$$S = \left\{ \begin{array}{l} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{array} \right\}$$

Let A and B be an event that we get a number 3 and 5 in the first and second rolled respectively, then we have:

$$A = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$$

So we have:

$$P(A)=\frac{6}{36}=\frac{1}{6}$$

And

$$B=\{(1,5),(2,5),(3,5),(4,5),(5,5),(6,5)\}$$

So we have:

$$P(B)=\frac{6}{36}=\frac{1}{6}$$

Also we get that:

$$A\cap B=\{(3,5)\}$$

So we have:

$$P(A\cap B)=\frac{1}{36}=\frac{1}{6}\cdot\frac{1}{6}=P(A)\cdot P(B)$$

This mean that the two events A and B are independent.

Multiplication rule 1 can be extended to three or more independent events by using the formula:

$$P(A\cap B\cap C\cap\cdots\cap Y)=P(A)\cdot P(B)\cdot\cdots\cdot P(Y)$$

The following example illustrated this extension

Example 5

Approximately 11% of men have a type of color blindness that prevents them from distinguishing between red and green colors. If three men were selected at random. Find the probability that all of them will have type of red-green color blindness.

Solution: : Let A , B and C an event that the first, second and third man has denoted to red-green blindness then the three events A, B and C are independent.

Therefore we get that:

$$P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C)$$

$$=(0.11)\cdot(0.11)\cdot(0.11)$$

$$=0.0001331\approx 0.00013$$

Example 6

If A and B are two independent events from the space of elementary events S, where $P(A) = 0.4$ and $P(B) = 0.6$. Find $P(A \cap B)$.

Solution: : Sine the two events are independent, by using the multiplication rule 1, we get

$$P(A \cap B) = P(A) \cdot P(B)$$
$$= (0.4)(0.6) = 0.24.$$

Example 7

Let $P(A) = 0.5, P(B) = 0.4, P(A \cup B) = 0.7$, where A and B are two events from the space of elementary events S. Verify whether A and B are independent events or not.

Solution: :

Step 1:
$$P(A \cup \mathrm{B}) = P(A) + P(B) - P(A \cap \mathrm{B})$$
$$0.7 = 0.5 + 0.4 - P(A \cap B)$$

Therefore
$$P(A \cap B) = 0.9 - 0.7 = 0.2.$$

Step 2:
$$P(A) \cdot P(B) = (0.5)(0.4) = 0.2.$$

From step 1 and step 2, we conclude
$$P(A \cap B) = P(A) \cdot P(B)$$

This mean that the two events A and B are independent.

In the previous examples with independent events, we replace the outcome to the space of elementary events then the number of the space of elementary events does not change. If we do not replace the outcome, the number of elements of the space of elementary events will decrease.

Example 8

A card is drawn from a deck and without replacement; then the second card is drawn.

a. Find the probability of getting a queen and then an ace.

b. Find the probability of getting a queen in the first and second draw.

Solution: :

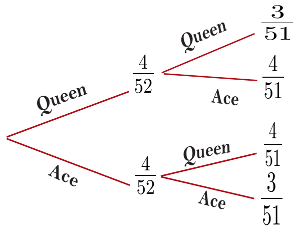
For (a): Let A and B be an event that we get a queen and an ace in the first and second drawn respectively, then we have:

$$P(\mathrm{A}) = \frac{4}{52} \text{ and } P(\mathrm{B}) = \frac{4}{51}.$$

Therefore we have:

$$P(\mathrm{A} \cap \mathrm{B}) = \frac{4}{52} \cdot \frac{4}{51} = \frac{4}{663} = P(\mathrm{A}) \cdot P(\mathrm{B}) = 0.006$$

Using the probability tree we get the following figure:



For (b): Let C be an event that we get a queen in the first and second drawn, then we have:

$$P(C) = \frac{4}{52} \cdot \frac{3}{51} = \frac{12}{2652} = \frac{1}{221} = 0.0045$$

Events are said to be independent of one another, if the occurrence of the first event in no way affect the outcome of the second event. On the other hand, when the occurrence of the first event changes the probability of the occurrence of the second event, the two events are said to be **dependent**.

Definition 4.3.2

If the occurrence of the first event affects the occurrence of the second event in such a way the probabilities are changed, the events are said to be dependent events.

Some examples of dependent events:

1. Draw a card from a deck without replacing it, and then draw a second card.
2. Selecting a ball from an urn without replacing it, and then selecting a second ball.

To find the probabilities when events are dependent, use the multiplication rule with a modification in notation. For the problem that discussed in example 8, the probability of getting an ace on the first draw is $\frac{4}{52}$, and the probability of getting a king on the second draw is $\frac{4}{51}$. By the multiplication rule, the probability of both events occurring is

$$P(\mathrm{C}) \cdot P(\mathrm{D}) = \frac{4}{52} \cdot \frac{4}{51} = \frac{16}{663} = 0.006$$

Where we had assumed that C and D be an event that we getting an ace and a king in the first and second drawn respectively.

B Conditional Probability

Example 9

An urn contains 5 red balls, 4 blue balls and 3 white balls, a ball is selected without replacement and it's color recorded. A second ball is selected and its color recorded. Find the probability that getting

a. The first ball is red and the second ball is white.

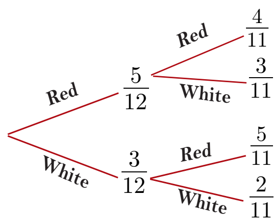
b. The first ball is red and the second ball is red.

Solution: : Let R and W be an event that we get a red, blue and white ball by selecting, then we have:

$$P(\mathrm{R}) = \frac{5}{12} \text{ and } P(\mathrm{W}) = \frac{3}{11}$$

$$P(\mathrm{R} \cap \mathrm{W}) = P(\mathrm{R}) \cdot P(\mathrm{W})$$

$$= \frac{5}{12} \cdot \frac{3}{11} = \frac{5}{44}$$



Also, we have

$$P(\mathrm{R} \cap \mathrm{R}) = P(\mathrm{R}) \cdot P(\mathrm{R})$$

$$= \frac{5}{12} \cdot \frac{4}{11} = \frac{5}{33}$$

The situation in calculating the probability in example 9 will differ when we are interested in calculating the probability of the second ball is white "given that" the first ball is red. This type of probabilities is called conditional probability. The conditional probability of an event B in relationship to an event A is defined as the probability that event B occurs after event A has already occurred. To find the conditional probability of an event B given

A denoted by $P(B|A)$ dividing both sides of the equation for multiplication rule 2 by $P(A)$.

Definition 4.3.3

If A and B are two events, then the probability of B given A which denoted by $P(B \mid A)$can be computed by the formula:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) \neq 0.$$

According to definition 4.3.3, then the probability that the event A occurs given that the event B is occurred and the probability that the event B occurs given that the event A is occurred are not equals. In mathematical notation

$$P(A \mid B) \neq P(B \mid A).$$

Result 2

If A and B are two independent events, then $P(B \mid A) = P(B)$ and $P(A \mid B) = P(A)$.

Let us consider the following example

Example 10

$P(A) = 0.5,\ P(B) = 0.6,\ P(A \cap B) = 0.2$. Find

a. $P(A \mid B)$ b. $P(B \mid A)$

Solution: :

a. $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.2}{0.6} = \dfrac{1}{3}$

b. $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{0.2}{0.5} = \dfrac{2}{5}$

Multiplication Rule 2

If A and B are two events, then the probability of both occurring is

$$P(A \cap B) = P(A) \cdot P(B \mid A)$$

or

$$P(A \cap B) = P(B) \cdot P(A \mid B)$$

The multiplication rule 2 can be illustrated in the following example

Example 11

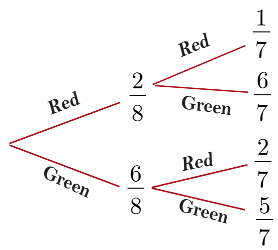
Eight toys are placed in a container. The toys are identical except for color, 2 are red and 6 are green. A child is asked to choose two toys at random. Find the probability that the child choose the two red toys.

Solution: : We can use a tree diagram:

let R_1 the event that get a Red in the first draw, and R_2 the event that get a Red in the second draw, then

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2 \mid R_1)$$

$$= \frac{2}{8} \cdot \frac{1}{7} = \frac{1}{28}$$



Example 12

Consider the experiment in which three coins are tossed. Let A be the event that the toss results at least one head. Find $P(A)$.

Solution: : Let $\bar{A}$ is the collection of simple events implying the event (three tails), then

$$\bar{A} = \{TTT\} \text{ and } P(\bar{A}) = P(\{TTT\}) = \frac{1}{8}$$

Using the definition of the complement event, then

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{1}{8} = \frac{7}{8}$$

Example 13

A record survey asked 100 people if they thought women in armed forces should be permitted to participate in combat. The results of the survey are shown in the following table:

Gender	Yes	No	Total
Male	40	10	50
Female	15	35	50
Total	**55**	**45**	**100**

Find the probabilities:

A. The respondent answered yes, given that the respondent is a female.
B. The respondent is a male, given that the respondent answered no.

Solution: : Let $M, F, Y,$ and N are events represent the response were male, female, answered yes, and answered no respectively. Then

A. Using result 2 with $P(Y \cap F) = 0.15$ and $P(F) = 0.5$, then

$$P(Y \mid F) = \frac{P(Y \cap F)}{P(F)} = \frac{0.15}{0.50} = \frac{15}{50} = 0.3$$

B. Using result 2 with $P(M \cap N) = 0.10$ and $P(N) = 0.45$, then

$$P(M \mid N) = \frac{P(M \cap N)}{P(N)} = \frac{0.10}{0.45} = \frac{10}{45}$$

Previous examples 11,12 and 13 discussed different cases of calculating probabilities including a condition. One more rule student should know is about probabilities is that the difference rule.

More rules in probability

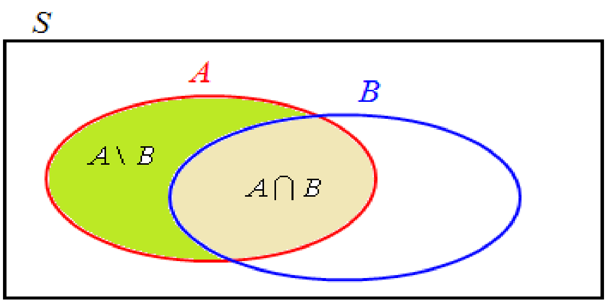
Another relationship for two events is the difference between them. The difference between events A and B denoted by $A \backslash B$ which is include all outcomes in the event A except the outcomes in the event B. Probability of such relationship is illustrated in the following result.

Result 3

If A and B are two events taken from the sample space S, then:

$$P(A \backslash B) = P(A) - P(A \cap B)$$

In order to ensure the validity of the previous result we can use the Venn diagrams as follows



According the figure above, the event A can be written as follows

$$A = (A\backslash B) \cup (A \cap B)$$

Taking the probability of both sides, we get

$$P(A) = \mathrm{P}[(A\backslash B) \cup (A \cap B)]$$

Using the fact that $A\backslash B$ and $A \cap B$ are mutually exclusive, then applying the addition rule 1, we get

$$P(A) = \mathrm{P}(A\backslash B) + \mathrm{P}(A \cap B)$$

Solving this equation for $\mathrm{P}(A\backslash B)$, then

$$\mathrm{P}(A\backslash B) = P(A) - \mathrm{P}(A \cap B)$$

Another notation can be used instead of $A\backslash B$ is $A \cap \overline{B}$

Example 14

If A and B are two events taken from the set of elementary events S, where

$P(A) = 0.7$, $P(B) = 0.4$,and $P(A \cap B) = 0.3$. Find:

a. $\mathrm{P}(A\backslash B)$ b. $\mathrm{P}(B\backslash A)$

Solution: : Applying the result, we get

a. $P(A \backslash B) = P(A) - P(A \cap B) = 0.7 - 0.3 = 0.4$

b. $P(B \backslash A) = P(B) - P(A \cap B) = 0.4 - 0.3 = 0.1$

Exercises 4.3

1 Let $P(A) = 0.5$, $P(B) = 0.4$, and $P(A \cap B) = 0.2$. Find:

a $P(A \mid B)$

b $P(B \mid A)$

c are A and B independent

d $P(A \setminus B)$

2 At a local university 52% of incoming first-year students have computers. If 3 students are selected at random. Find the following probabilities:

a None have computers.

b All have computers.

3 When 2 cards are selected from a standard deck of 52 cards without replacement, find the following probabilities:

a Both are diamond.

b Both are the same color.

c Both are aces.

4 It is reported that 72% of working women use computers at work, choose 4 working women at random. Find the probability that:

a All 4 use a computer at their work.

b At least one does not use a computer at work.

5 A flashlight has 8 batteries, 3 at which are defective. If 2 are selected at random without replacement, find the probability that both are defective.

6 40% of people who own cell phones use their phones to send and receive text messages. Choose 5 cell phones owners at random; find the probability that none use their phones for texting messages.

7 In a department store there are 150 customers. 100 of whom who will buy at least one item. If 3 customers are selected at random one by one, find the probability that all will buy at least one item.

8 Three cards are drawn from a deck without replacement. Find the probabilities:

a All are black cards. b. All are spades. c. All are kings.

9 An urn contains 6 white balls, 5 red balls and 4 green balls. 3 balls are selected at random without replacement. Find the probability that all are from the same color.

10 The medal distribution from 2004 summer Olympic Games for top 23 countries is shown below:

Country	Gold	Silver	Bronze
U.S.	35	39	29
Russia	27	27	38
China	32	17	14
Australia	17	16	16
Others	133	136	153

Choose one medal winner at random:

a Find the probability that the winner won the gold medal, given that the winner was from China.

b Find the probability that the winner was from Australia, given that he or she won a silver medal.

11 The weekly salaries (in dollars) of sixteen government workers are listed below. Find the first quartile (Q_1).

	Cookies	Mugs	Candy
Coffee	20	13	10
Tea	12	10	12

Choose one basket at random. Find the probability that it contains:

a Coffee or candy.

b Tea given that it contains mugs.

12 The following table represents theRh factor in blood types

	O	A	B	AB
Rh^+	37%	34%	10%	4%
Rh^-	6%	6%	2%	1%

Choose one American at random. Find the probability that the person has type A given that he is Rh^-.

13 Of Ph.D. Students, 60% have paid assistantships. If 3 of them are selected at random, find the probability that:

a All have assistantship.

b None have assistantship.

c At least one has an assistantship.

14 A coin is tossed 4 times. Find the probability of getting at least one tail.

15 A die is rolled 3 times. Find the probability of getting at least one even number.

16 It was reported that 20% of computer games sold in 2010 were classified as "family and children's". Choose 5 purchased computer games at random. Find the probability that:

a None of 5 was family and children's.

b At least 1 of 5 was family and children's.

17 If 4 cards are drawn from a deck of 52 cards and not replaced. Find the probability of getting at least one spade.

18 If A and B are two events taken from the set of elementary events S, such that $P(A)$=0.6 and $P(A \cap B) = 0.3$, then any of the following values (A, B, C or D) equal to $P(A \setminus B)$?

A) 0.3 B) 0.03 C) 0.5 D) 0.05

19 If $P(A)$=0.4, $P(B)$=0.1 $P(A \cap B) = 0.2$, then any of the following values (A, B, C or D) equal to $P(B \mid A)$?

A) 0.9 B) 0.09 C) 0.5 D) 0.05

20 If A and B are two independent events taken from the set of elementary events S, such that $P(A)$=0.4 and $P(B)$=0.1, then any of the following values (A, B, C or D) equal to $P(A \cap B)$?

A) 0.04 B) 0.4 C) 0.1 D) 0.004

21 If A and B are two independent events taken from the set of elementary events S, such that $P(A)= P(B)$, and $P(A \cap B)$ =0.09, then any of the following values (A, B, C or D) equal to $P(A \cup B)$?

A) 0.50 B) 0.51 C) 0.9 D) 0.001

4.4 Bayes' Rule

In this section we will find a probability of event using the well -known Bayes theorem. The Bayes theorem is named for well known Thomas Bayes (1702-1761), and is also referred as Bayes rule (or Bayes formula). The Bayes theorem becomes more commonly used in statistical inference. It is also a method for interpreting the evidence of previous experiment or it is a simple method to calculate the conditional probability when we have a partition of the certain event S.

Definition 4.4.1

Given $E_1, E_2, E_3, \cdots, E_n$ that are mutually exclusive events in a space of elementary events S, then $E_1, E_2, E_3, \cdots, E_n$ are a partition of S if the following condition is realized:

1) $E_i \neq \phi$ for all $i = 1,2,3,\ldots\ldots,n$

2) $\bigcup_{i=1}^{n} E_i = S$

Total Probability Formula

If we have $E_1, E_2, E_3, \cdots, E_n$ a partition of S, and let A an event in S, then the probability of the event A can be expressed as:

$$P(A) = P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2) + \cdots + P(E_n)P(A \mid E_n)$$
$$= \sum_{k=1}^{n} P(E_k)P(A \mid E_k).$$

The above equation is called total probability formula.
Let us consider the following example

Example 1

Let $\{E_1, E_2\}$ be a partition of the space of elementary events S with:

$$P(E_1) = 0.7 \text{ and } P(E_2) = 0.3$$

And assuming that for an event A from S we have:

$$P(A \mid E_1) = 0.2 \text{ and } P(A \mid E_2) = 0.3$$

Then calculate $P(A)$.

Solution: Using the total probability formula we get:

$$P(A) = P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)$$
$$= 0.7 \times 0.2 + 0.3 \times 0.3 = 0.23$$

The solution of example 1 can be generalized to the following rule.

Bayes' Theorem

Let $E_1, E_2, E_3, \cdots, E_n$ } be a partition of the space of elementary events S, and A an event from S. Then we have for any i=1,2,3,....,n the probability of E_i given A equal to

$$P(E_i \mid A) = \frac{P(E_i)\,P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i)\,P(A \mid E_i)}$$

The above equation called the Bayes formula in the probability. To understand this, consider example 2.

Example 2

If an experiment is conducted, one and only one of three mutually exclusive events E_1, E_2, and E_3 can occur with $P(E_1) = 0.2$, $P(E_2) = 0.5$, and $P(E_3) = 0.3$ and the probability of a fourth event A occurring, given that E_1, E_2, or E_3 occurs are:

$P(A \mid E_1) = 0.2,\ P(A \mid E_2) = 0.1,\ P(A \mid E_3) = 0.3.$

If A is observed. Find:

a. $P(E_1 \mid A)$ b. $P(E_2 \mid A)$ c. $P(E_3 \mid A)$

Solution:

$$\begin{aligned} P(A) &= P(E_1)\cdot P(A \mid E_1) + P(E_2)\cdot P(A \mid E_2) + P(E_3)\cdot P(A \mid E_3) \\ &= (0.2)(0.2) + (0.5)(0.1) + (0.3)(0.3) \\ &= 0.04 + 0.05 + 0.09 = 0.18 \end{aligned}$$

a. Applying Bayes theorem, we get

$$P(E_1 \mid A) = \frac{P(E_1)\,P(A \mid E_1)}{P(A)} = \frac{0.04}{0.18} = \frac{2}{9}$$

b. Applying Bayes theorem, we get

$$P(E_2 \mid A) = \frac{P(E_2)\,P(A \mid E_2)}{P(A)} = \frac{0.05}{0.18} = \frac{5}{18}$$

c. Applying Bayes theorem, we get

$$P(E_3 \mid A) = \frac{P(E_3)\,P(A \mid E_3)}{P(A)} = \frac{0.09}{0.18} = \frac{1}{2}$$

Example 3

In a factory for the production of light lamps there are three lines of production that these lines produce 50%, 30% and 20% respectively of the total production of the factory. The ratio of defected bulbs in the production of these lines is 0.03 and 0.02 and 0.01 respectively, and required the following:

a) We withdraw a lamp of the total production of the factory, what is the probability that this lamb is defected?

b) If we have found that the drawn lamp is defected, what is the probability that this drawn lamp from the first line of factory?

Solution:

For a) To answer these questions, we will assume that's E_1, E_2, E_3 the drawn lamp from the first, second and third production line respectively, we find that these events E_1, E_2, E_3 are a partition of the space of elementary events S (S is the total products of factory). Then we have:

$$P(E_1) = \frac{50}{100},\ P(E_2) = \frac{30}{100},\ P(E_3) = \frac{20}{100}$$

Now let A be an event that the drawn lamp is defected. So we get by using the total probability formula:

$$P(A) = \sum_{n=1}^{3} P(E_n) \cdot P(A \mid E_n) = \frac{50}{100} \times \frac{3}{100} + \frac{30}{100} \times \frac{2}{100} + \frac{20}{100} \times \frac{1}{100}$$
$$= 0.023$$

For b) By using the Bayes formula we get that required probability equal to:

$$P(E_1 \mid A) = \frac{P(E_1)P(A \mid E_1)}{\sum_{k=1}^{3} P(E_k).P(A \mid E_k)} = \frac{(0.50)(0.03)}{0.023} = 0.6522$$

Exercises 4.4

1 Medical case histories indicate the different illnesses may produce identical symptoms. Suppose a particular set of symptoms, which we will denote as event H, occurs only when any of three illnesses A, B or C occurs. (Assume and) are mutually exclusive. Studies show these probabilities of getting the three illnesses

$$P(A) = 0.01,\ P(B) = 0.005,\ P(C) = 0.02.$$

The probabilities of developing the symptoms H, given a specific illness are

$$P(H \mid A) = 0.9,\ P(H \mid B) = 0.95,\ P(H \mid C) = 0.75.$$

Assuming that an ill person shows symptoms H. Find the probability that the person has illness A. i.e. find $P(A \mid H)$

2 A space of elementary events S can be divided into two events E_1, E_2 that occur with probabilities 60% and 40% respectively. An event A occurs 30% of the time in the first event E_1 and 50% of the time in the second event E_2. What is the unconditional probability of the event A, regardless of which event it comes from.

3 There are two boxes, the first one contains 4 white balls, and 6 black balls and the second one contains 8 white balls and 3 black balls. If we choose one box at random and take a ball from it randomly, find the probability that:

a The chosen ball is black.

b If the chosen ball was black find the probability of getting it from the first box.

4 A factory has three machines I, II, III. If machine I produces 20% of the items, machine II produces 30%, and machine III produces 50% of the items, with defective from the machine as 5%, 2% and 1% respectively. If an item is selected at random, find the probability that the item is not defective.

5 A manufacturer makes two models of an item, model I, which accounts for 80% of units sales and model II, which accounts for 20% of units sales. Because of defects, the manufacturer has to replace 10% of its model I and 18% of its modelII. If a model is selected at random. Find the probability that it will be defective.

6 In a factory for the production of switches there are four lines of production so that produces these lines 25%, 20%%, 35% and 20% respectively of the total production of the factory.

The ratio of defected switches in the production of these lines is 0.01 , 0.02, 0.03 and 0.04 respectively, and required the following:

a We draw a switch of the total production of the factory, what is the probability that this switch is not defected ?

b If we have found that the drawn switch is defected, what is the probability that this drawn switch from the fourth line of factory ?

7 In a factory for the production of transistors there are five lines of production so that produces these lines are equal to gather. The ratio of defected switches in the production of these lines is 0.03 , 0.04, 0.01, 0.02 and 0.01 respectively, and required the following:

a We draw a transistor of the total production of the factory, what is the probability that this transistor is defected ?

b If we have found that the drawn transistor is not defected, what is the probability that this drawn transistor from the first line of factory ?

4.5 Counting Rules

Sometimes we are interested to know the number of all possible outcomes for a sequence of events. Three rules can be used to determine this number. Such rules are, the **fundamental counting rule**, the **permutation rule**, and the **combination rule**. First we will explain these rules and they will be used latter to find probabilities. Moreover we will learn about factorials, combinations, and permutations.

A The Fundamental Counting Rule

In many experiments the number of outcomes is very large, and it is not easy to list all outcomes. In such cases, we may use the counting rule to find the total number of outcomes.

Fundamental Counting Rule

In a sequence of n events in which the first one has k_1 possibilities, the second event has k_2 and the third has k_3 and so forth, the total number of possibilities of the sequence will be:

$$k_1 \cdot k_2 \cdot \cdots \cdot k_n$$

The following examples are illustrating the fundamental counting rule.

Example 1

A coin is tossed and a die is rolled. Find the number of outcomes for the sequence of events.

Solution: A coin has 2 faces and a die has 6 faces, then the number of possible outcomes is:

$$(2).(6)=12$$

Example 2

A paint manufacturer wishes to manufacture several different paints. The categories include:
Color: Red, blue, white, black, green, brown, yellow.
Type: Latex, oil.
Texture: Flat, semi-gloss, high gloss.
Use: outdoor, indoor.
How many different kind of paint can be made if you can select one color, one type, one texture, and one use.

Solution: Color: 7 Type: 2 Texture: 3 Use: 2. The possibility is 7(2)(3).(2)=84 outcomes.

Example 3

The manager of a department store chain wishes to make four-digit identification cards for his employees. How many different cards can be made if he uses the digits 1,2,3, 4 and 5 and repetitions are permitted?

Solution: The number of cards is (5) (5) (5) (5)= 625. Now, in the case the repetitions are not permitted. Then, the first digit can be chosen in 5 ways, the second digit can be chosen in 4 ways, the third digit can be chosen in 3 ways, and the fourth digit can be chosen in 2 ways. So, the number of possibilities is: (5) (4) (3) (2)= 120

Factorial Notation

The factorials denote by " ! " (read as factorial). The value of the factorial of a number is obtained by multiplying all the integers from that number to 1.

Definition 4.5.1

The number of ways of arranging n distinct object is

$$n! = n(n-1)(n-2)\cdot \cdots \cdot 2\cdot 1$$

with $0! = 1$

Let us consider an example

Example 4

Evaluate each of the following:

a. $4!$ b. $7!$ c. $(13-8)!$

Solution:

a. To evaluate 4!, we multiply all the integers from 4 to 1. Therefore

$$4! = 4\cdot 3\cdot 2\cdot 1 = 24$$

b. Similarly, we can evaluate 7!

$$7! = 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1 = 5040$$

b. The value of $(13-8)!$ is the same as 5!. Therefore

$$(13-8)! = 5! = 5\cdot 4\cdot 3\cdot 2\cdot 1 = 120$$

The second rule of counting is the permutations

B Permutations

The arrangement of n objects in a specific order r objects at a time is called a permutation of n objects taking r objects at a time. In simple words, we may define the permutation as follows

Definition 4.5.2

A permutation is any arrangement of n objects in specific order.

Example 5

Suppose a business owner has a choice of 5 locations in which to establish his business. He decides to rank only top 3 of the 5 locations. How many different possibilities can he rank.

Solution: The first choice is 5, the second choice is 4, and the third choice is 3. Therefore we have the possibilities equal to: $5 \cdot 4 \cdot 3 = 60$

Permutation Rule

The arrangement of n objects in a specific order r objects at a time is called a permutation of n objects taking r objects at a time. It is denoted by $n\mathrm{P}r$ and cab be calculated as follows:

$$n\mathrm{P}r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

Example 6

Evaluate each permutation

$$5\mathrm{P}3, 5\mathrm{P}4, 5\mathrm{P}5, \text{ and } 6\mathrm{P}2$$

Solution: Applying the permutation rule, then

$$5\mathrm{P}3 = 5 \cdot 4 \cdot 3 = 60$$

$$5\mathrm{P}4 = 5 \cdot 4 \cdot 3 \cdot 2 = 120$$

$$5\mathrm{P}5 = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

$$6\mathrm{P}2 = 6 \cdot 5 = 30$$

Example 7

A television news director wishes to use 3 news stories of an evening show. One story will be the lead story, one will be the second story, and the last will be a closing story. If he has a total of 10 stories to choose from. How many possible ways can the program are set up.

Solution: Since the order is important, so it is permutation and hence

$$10P3 = \frac{10!}{(10-3)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{7!} = 720$$

The third rule of counting is the permutations.

C Combination

Suppose a book designer wishes to select two colors of material to design a statistical book, and he has four colors. How many different possibilities can there be in this situation?

This problem differs from previous ones in that the order of selection is not important. That is, if the designer selects yellow and red, this selection is the same as the selection red and yellow. This type of selection is called a **combination**. The difference between a permutation and a combination is that in a combination, the order or arrangement of the objects is not important, while order is important in a permutation.

Definition 4.5.3

A selection of distinct objects without regard to order is called a **combination**.

Example 8,9, and 10 are illustrates the difference between a permutation and a combination.

Example 8

Given the letters A,B and C, list the permutations and combinations for selecting 2 letters.

Solution: The permutations are:

$$AB,\ AC,\ BC,\ BA,\ CA,\ CB$$

and the combinations are: AB , AC, BC.

Combination Rule

The number of combinations of r objects selected from n objects is denoted by $n\mathrm{C}r$ given by the formula:

$$n\mathrm{C}r = \frac{n!}{r!(n-r)!}$$

Example 8

How many combinations of 6 objects are there, taken 3 at a time.

Solution: Since the order is not important, the combination is:

$$6C3 = \frac{6!}{3!(6-3)!} = \frac{6!}{3! \cdot 3!} = \frac{6 \cdot 5 \cdot 4 \cdot 3!}{3! \cdot 3!} = 20$$

Note that the relation between permutations and combinations is:

$$nCr = \frac{nPr}{r!}$$

Example 9

In a hospital, there are 8 nurses and 4 doctors. A committee of 3 nurses and 2 doctors is to be chosen. How many different possibilities are there?

Solution:

There are $8C3$ possibilities to choose the nurse, that is

$$8C3 = \frac{8!}{3! \cdot 5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{(3 \cdot 2 \cdot 1) \cdot 5!} = 56$$

There are $4C2$ possibilities to choose the doctors, that is

$$4C2 = \frac{4!}{2! \cdot 2!} = \frac{4 \cdot 3 \cdot 2!}{2! \cdot 2!} = 6$$

There are $(8C3)(4C2)$ possibilities to choose the committee

$$8C3 \cdot 4C2 = 56(6) = 336$$

At the end of this chapter, the student now is capable of finding the probabilities in any probability experiment whether the number of outcomes is small or large, in addition to calculate probabilities for two or more events using the addition and multiplication rules, chapter 5 discusses step forward in finding the probability distribution for variables.

Exercises 4.5

1 How many 6-digits zip codes are possible if digits can be repeated? If digits cannot be repeated.

2 How many different ways can 5 different books be arranged on a shelf.

3 In how many ways can 7 students be seated in a row on a stage.

4 Evaluate each of these:

a	$7!$	b	$8!$
c	$4P3$	d	$10P4$
e	$5P5$	f	$nP0$
g	$6C2$	h	nCn

5 There are 4 major roads from city A to B and 3 major roads from B to C. How many different trips can be made from A to C passing through city B.

6 How many different ID cards can be made if there are 6 digits on a card and no digit can be used more than once.

7 How many ways can 5 tickets are selected from 40 tickets if each ticket wins a different prize.

8 How many ways can an adviser choose 5 students from a class of 20 if they are all assigned the same task? How many ways can the students be chosen if they are each given a different task.

9 An investigative agency has 8 cases and 4 agents. How many different ways the cases be assigned if only one case assigned to each agent.

10 How many different 3 letter permutations can be formed from the letters in the word "UNIVERSAL".

11 How many ways can a jury of 5 women and 5 men be selected from 10 women and 15 men.

12 There are 15 seniors and 20 juniors in a particular social organization. In how many ways can 4 seniors and 3 juniors be chosen to participate in a charity event.

13 How many ways can a person select 6 televisions commercials from 12 television commercials.

14 How many ways can a person select 5 DVDs from a display of 12 DVDs.

15 How many ways can a buyer select 7 different posters from 20 posters.

16 How many different ways can an instructor select 3 textbooks from 15.

17 How many ways can a person select 6 videotapes from 11 videotapes.

18 How many different signals can be made by using at least 3 different flags if there are 7 different flags from which to select.

19 How many ways can a dinner Parton select 4 appetizers and 3 vegetables if there are 7 appetizers and 5 vegetables on the menu.

20 In a board of directors composed of 7 people. How many ways can one chief executive office, one directors, and one treasurer be selected.

21 Any of the following values (A, B, C or D) equal to the value 5! ?

A) 100 B) 20 C) 25 D) 120

22 Any of the following values (A, B, C or D) equal to the value $7P2$?

A) 14 B) 42 C) 2 D) 50

23 Any of the following values (A, B, C or D) equal to the value 9C3 ?

A) 6 B) -6 C) 84 D) 27

24 In a hospital, there are 10 nurses and 3 doctors. A committee of 4 nurses and 2 doctors is to be chosen. How many different possibilities are there?

A) 60 B) 210 C) 840 D) 630

4.6 Probability and Counting Rules

In many types of probability problems, the counting rules can be combined with the probability rules to find probabilities. By using the fundamental counting rule, the permutation rules, and the combination rule, we can compute the probability of outcomes of many experiments.

Example 1

A box contains 12 transistors, 3 of which are defective. If 5 are taken at random. Find the following probabilities:

a Exactly two are defectives.

b None of them is defective.

c At least one is defective.

Solution: There are 12C5ways to choose 5 transistors, and 12C5 stands as a number of the sample space which is the denominator in each case.

Now let be:

A an event that we have exactly 2 defectives,

B an event that we have no defectives,

C an event that we have at least 1 defectives,

Then for:

a. Two defective transistors can be selected as 3C2and the other three non-defective transistors can be selected as 12C3. Therefore

$$P(A)=\frac{(3C2)(9C3)}{(12C5)}=\frac{5}{6}$$

b. $P(B)=\frac{(9C5)}{12C5}=\frac{7}{44}$

c. $P(C)=1-P(B)=1-\frac{7}{44}=\frac{37}{44}$

Example 2

A combination lock consists of the 26 letters of the alphabet. If a 3-letter combination is needed. Find the probability that the combination will consist the letters ABC in that order. The same letter can be used more than once.

Solution: Since repetitions are permitted, then there are (26)(26)(26)=17576 different possible ways. And since there is only one ABC combination,

$$P(ABC) = \frac{1}{(26)^3} = \frac{1}{17576}$$

Exercises 4.6

1 A committee consisting of 4 people is to be formed from 15 parents and 6 teachers. Find the probability that the committee will consists of these people:

a All teachers.

b 2 parents and 2 teachers.

2 In a company there are 7 executives: 4 women and 3 men. Three are selected to attend a seminar. Find the probability that:

a All 3 selected will be women.

b All 3 selected will be men.

c 2 men and 1 woman will be selected.

d 1 man and 2 women will be selected.

3 A package contains 10 resistors, 2 of which are defective. If 4 are selected, find the probability of getting:

a 0 defective resistors.

b 1 defective resistors.

c 2 defective resistors.

4 When 3 dice are rolled. Find the probability of getting a sum of 7.

5 A student prepares for an exam by studying a list of 10 problems. He can solve 6 of them. For the exam, the instructor selects five questions at random from the list of 10. Find the probability that the student can solve all five problems on the exam.

6 A family has 5 children. Find the probability that:

a All boys.

b All boys or girls.

c Exactly 2 boys or 2 girls.

d At least one child of each gender.

7 Roll two dice and multiply the numbers. Find the probability that the product is a multiple of 6.

8 A coin tossed 3 times. Find the probability of getting at least one head.

9 Three dice are rolled. Find the probability of getting triples.

10 From 25 students we want to select 3 of them to make a committee where the first is a leader, the second and the third are members.

11 A bag contains 6 red marbles,3 blue marbles, and 7 green marbles. If a marble is randomly selected from the bag, what is the probability that it is blue?

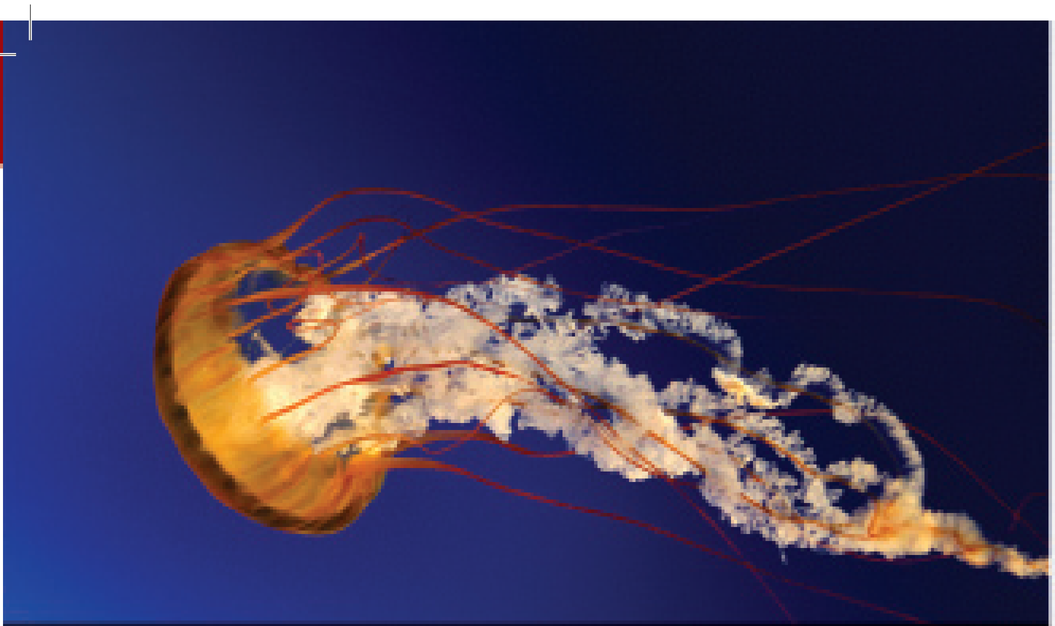
A) 1/7 B) 1/6 C) 1/3 D) 3/16

12 Six people sitting on six seats in a row, any of the following values (A, B, C or D) equals the number of sitting possibilities?

A)) 6C6 B) 6P6 C) 6P1 D) 6C1

13 Six people sitting on six seats on a round circle seats, any of the following values (A, B, C or D) equals the number of sitting possibilities?

A) 6C1 B) 6P6 C) 5P5 D) 6C5



CHAPTER

5

Random Variables and Their Probability Distributions

OBJECTIVES

1. Know what random variable is.
2. Know what a discrete random variable is.
3. Know what a continuous random variable is.
4. Construct a probability distribution for a random variable.
5. Determine the mean, variance, standard deviation, and expected value for a discrete random variable.
6. Find the exact probability for x-successes in n trials of a binomial experiment.
7. Calculate the mean, variance, and standard deviation for the variable of a binomial distribution.
8. Identify a continuous distribution.
9. Identify the properties of a normal distribution.
10. Find the area under the standard normal distribution, given various z-values.
11. Find probabilities for a normally distributed variable by transforming it into standard normal distribution.
12. Find specific data values for given percentages, using the standard normal distribution.

Random Variables and their Probability Distributions

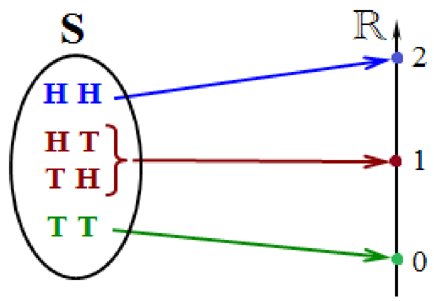
In this chapter, random variables and its type are explained. Moreover, some of main concepts that related to random variables such as probabilities distribution, mean and standard deviation are discussed. And at the end of this chapter, special probability distributions the binomial and normal probability distributions are illustrated.

5.1 Random variables

A coin is tossing two times, then the space of elementary events S contains four results $w_1=HH$, $w_2=HT$, $w_3=TH$, $w_4=TT$. Now we suppose that a real function X defined on the space of elementary events S as follow:

$$X : S = \{\underbrace{HH}_{w_1}, \underbrace{HT}_{w_2}, \underbrace{TH}_{w_3}, \underbrace{TT}_{w_4}\} \longrightarrow \mathrm{R}$$

$$w \mapsto X(w) = \begin{cases} 0 & \text{for } w = TT \\ 1 & \text{for } w = TH, HT \\ 2 & \text{for } w = HH \end{cases}$$



So we note that this function has three real values 0, 1 and 2. and these are random values because the results $w_1=HH$, $w_2=HT$, $w_3=TH$, $w_4=TT$ are random results. Therefore we can consider to the set $E=\{0,1,2\}$ as the results of a random experiment also. Now, if the inverse image of any event B of E is an event of S, then X is said to be a random variable on S.

Refer to the previous example, we note that:

$X^{-1}(\{0\}) = \{TT\}$ is an event of S

$X^{-1}(\{1\}) = \{HT, TH\}$ is an event of S

$X^{-1}(\{2\}) = \{HH\}$ is an event of S

So the function X is a random variable on S.

Also we can define the random variable as follow.

Definition 5.1.1

A function X defined on a space of elementary events S with values in a set $E \subseteq \mathrm{R}$ is a random variable on S if and only if the inverse image of any event B of E is an event of S.

Random variables are classified as either discrete or continuous, according to the assumed values of X. The distinguish between discrete and continuous random variables is very important in probability theory because different techniques are used to describe their distributions.

A Discrete Random Variable

A random variable that assumes values can be counted is called discrete random variable.

Definition 5.1.2

A random variable that assumes countable values is called a discrete random variable.

Here are some examples of discrete random variables:
1. X describes the number of tails obtained in tossing a coin twice.
2. X describes the number of complaints received at the office of an airline on a given day
3. X describes the number of cars sold during a given week
4. X describes the number of customers who visit a bank during any given day

B Continuous Random Variable

A continuous random variable is a random variable has uncountable values. A continuous random variable can assume any value over an interval or more then one.

Here are some examples of continuous random variables:
1. X describes the remaining time to end a certain test
2. X describes the time taken to travel from a city to another
3. X describes the weight of a travel bag
4. X describes the temperatures recorded at a certain hour during a day of the week

Exercises 5.1

1. Classify each of the following random variables as discrete or continuous.
   a. The number of cars sold during a given month
   b. The price of a shirt
   c. The number of customers who visit a hospital during any given hour
   d. The number of accidents that recorded during a certain day
2. Give some examples on a discrete random variables
3. Give some examples on a continuous random variables

5.2 Probability Distribution of a Discrete Random Variable

In general, the probability distribution of a discrete random variable describes how the probabilities are distributed over all the outcomes of that random variable.

Definition 5.2.1

The probability distribution for a discrete random variable is a formula, table or graph that provides $P(X{=}x)$, the probability associated with each of the values x of X.

The events associated with different values of X cannot overlap because one and only one value of X is assigned to each elementary event; hence, the values of x represent mutually exclusive numerical events. Sum in $P(X{=}x)$ overall values x equals the sum of the probabilities of all elementary events and hence equals to one. Moreover, for a discrete random variable the probability distribution possesses two characteristics (conditions) and these two characteristics must satisfy in any probability distribution

Requirements for a Discrete Probability Distribution

The probability distribution of a discrete random variable possesses the following two characteristics. For each value of the random variable X:

1. $0 \le P(X = x) \le 1$ **2.** $\sum_x P(X = x) = 1$

The following examples illustrate the concept of the probability distribution of a discrete random variable.

Example 1

Each of the following tables lists certain values of X and their probabilities. Determine whether or not each table represents a valid probability distribution.

a.

x	$P(X{=}x)$
0	0.01
1	0.05
2	0.30
3	0.20

b.

x	$P(X{=}x)$
0	0.50
1	-0.20
2	-0.05
3	0.75

c.

x	$P(X{=}x)$
0	0.00
1	0.77
2	0.23

Solution:

(a) The first condition of a probability distribution is satisfied, this because each probability listed in this table is in the range 0 to 1. However, the second condition of a probability distribution is not satisfied, because the sum of all probabilities is not equal to 1.0 .Therefore, this table does not represent a valid probability distribution.

(b) The sum of all probabilities listed in this table is equal to 1.0, but two of the probabilities are negative. This violates the first condition of a probability distribution. Therefore, this table does not represent a valid probability distribution.

(c) Each probability listed in this table is in the range 0 to 1. Also, the sum of all probabilities is equal to 1.0 . Therefore, this table represents a valid probability distribution.

Example 2

The following table lists the probability distribution of the number of defectives per hour for a machine production.

Defectives per hour	0	1	2	3	4
Probability	0.33	0.25	0.17	0.15	0.10

Find the probability that the number of defectives for this machine production through a given hour is

a. Exactly 2 b. 1 to 3 c. Less than 3 d. At least 2

Solution:

Let x denote the number of defectives for this machine production through a given hour. According to the given table, we can calculate the required probabilities as follows.

a. The probability of exactly two defectives

$$P(\text{exactly 2 defectives}) = P(X=2) = 0.17$$

b. The probability of 1 to 3 defectives is given by the sum of the probabilities of 1, 2, and 3 defectives:

$$\begin{aligned} P(1 \text{ to } 3 \text{ defectives}) &= P(1 \le X \le 3) \\ &= P(X=1) + P(X=2) + P(X=3) \\ &= 0.25 + 0.17 + 0.15 = 0.57 \end{aligned}$$

c. The probability of less than 3 defectives is obtained by adding the probabilities of 0, 1, and 2 defectives:

$$P(less\ \text{than } 3 \text{ defectives}) = P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0.33 + 0.25 + 0.17 = 0.75$$

d. The probability of at least 2 defectives is given by the sum of the probabilities of 2, 3, and 4 defectives:

$$P(at\ \text{least } 2 \text{ defectives}) = P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$$
$$= 0.17 + 0.15 + 0.10 = 0.42$$

Example 3

Suppose that we toss two fair coins once, and let X represents the number of heads observed. Find the probability corresponds for the possible outcomes.

Solution: The outcomes for such experiment can be organized as in the following table

Simple Event	Coin 1	Coin 2	$P(E_i)$	x
E_1	H	H	1/4	2
E_2	H	T	1/4	1
E_3	T	H	1/4	1
E_4	T	T	1/4	0

Now, according to the constructed table, we can compute the probability values for each outcome as follows:

$$P(X = 0) = P(E_4) = \frac{1}{4}$$

$$P(X = 1) = P(E_2) + P(E_3) = \frac{1}{2}$$

$$P(X = 2) = P(E_1) = \frac{1}{4}$$

And it can be represented as in the following table

x	0	1	2
$P(X=x)$	1/4	2/4	1/4

Exercises 5.2

1 Each of the following tables lists certain values of X and their probabilities. Verify whether or not each represents a valid probability distribution and explain why.

a.

x	$P(X=x)$
2	-.20
3	0.10
4	0.20
5	0.10

b.

x	$P(X=x)$
2	0. 50
3	0.20
4	0.10
5	0.20

c.

x	$P(X=x)$
7	0.00
8	0.77
9	0.40

d.

x	$P(X=x)$
1	.20
2	0.10
3	0.20
4	0.60

e.

x	$P(X=x)$
0	0. 30
1	0.20
2	0.10
3	0.20

f.

x	$P(X=x)$
5	0.00
6	-0.50
7	0.10

2 The following table gives the probability distribution of a discrete random variable X

x	0	1	2	3	4	5
$P(X=x)$	0.20	0.15	0.13	0.02	0.10	0.40

Find the following probabilities

a $P(X = 5)$

b $P(X \leq 1)$

c $P(2 \leq X \leq 5)$

d $P(X > 5)$

e Probability that X assumes a value less than 2

f Probability that X assumes a value in the closed interval 0 to 4

g Probability that X assumes a value greater than 4

3 The following table gives the probability distribution of a discrete random variable X.

x	1	2	3	4	5
$P(X=x)$	0.23	0.33	0.13	0.10	0.21

Find the following probabilities

a $P(X = 0)$

b $P(X > 3)$

c $P(1 \le X \le 3)$

d $P(X \le 2)$

e Probability that X assumes a value less than 3

f Probability that X assumes a value in the interval 2 to 5

g Probability that X assumes a value greater than or equal 1

4 Suppose that two balanced dice are rolled. Let X be a random variable denotes the absolute value of the difference of the two numbers. What are the possible values of the random variable X ?

A) 0, 1, 2, 3, 4, 5 B) -5,-4,-3,-2,-1, 0, 1, 2, 3, 4, 5

C) 1, 2, 3, 4, 5 D) 0, 1, 2, 3, 4, 5, 6

5 The following table gives the probability distribution of a discrete random variable X

x	0	1	2	3	4	5
$P(X=x)$	0.15	0.15	0.13	0.2	0.10	0.27

Then , then any of the following values (A, B, C or D) equal to $P(X<3)$?

A) 2 B) 0.43 C) 0.3 D) 0.57

6 The following table gives the probability distribution of a discrete random variable X

x	0	1	2	3	4
$P(X=x)$	0.34		0.13	0.10	0.10

Then the missing value is

A) -1 B) 0.33 C) 0.73 D) 0.6

5.3 The Mean and Standard Deviation for a Discrete Random Variable

In this section, we will learn how to calculate the mean and the standard deviation of a discrete random variable.

A Mean of a Discrete Random Variable

The mean (or expected value) of a discrete random variable is the value that we expect to observe per repetition, on average, if we perform an experiment for a large number of times. Consider the following example; assume X to be the number of heads observed during tossing a coin twice. Then the following probability distribution table could be constructed to represent the results of our experiment.

x	0	1	2
$P(X=x)$	1/4	2/4	1/4

Suppose the experiment is repeated for a large number of times, for instance 100 times intuitively, you would expect to observe approximately one hundred zeros, 2 hundred ones, and one hundred twos, then the average value of X would equal

$$\text{The average value of } X = \frac{100(0)+(200)(1)+(100)(2)}{400}$$
$$= \frac{100}{400}(0)+\frac{(200)}{400}(1)+\frac{(100)}{400}(2)$$
$$= \frac{1}{4}(0)+\frac{(1)}{2}(1)+\frac{(1)}{4}(2) = 1$$

Note that,the first term in this sum is $0.(P(X=0))$, the second term is $1.(P(X=1))$ and the third is $2.(P(X=2))$. The average value of X can be formulated as $\sum_x x\,P(X=x)=1$.

Definition 5.3.1

Let X be a discrete random variable with probability distribution $P(X=x)$. The mean or expected value of is given as

$$\mu = \mathrm{E}(X) = \sum_x x \cdot P(X=x)$$

Where x is the elements are summed over all values of the random variable X.

Example 1 illustrates the calculation of the mean of a discrete random variable

Example 1

The following table gives the probability distribution for the price of a shirt per week.

Price of a shirt	1	2	3	4	5
Probabilities	0.10	0.50	0.20	0.15	0.05

Find the mean price of a shirt per week.

Solution: To find the mean price of a shirt per week, we multiply each value of the price of the shirt x by its probability $P(X=x)$ and then add these products. The products $xP(X=x)$ are listed in the third column of the following Table.

x	$P(X=x)$	$xP(X=x)$
1	0.10	$1(0.10)=0.10$
2	0.50	$2(0.50)=1.00$
3	0.20	$3(0.20)=0.60$
4	0.15	$4(0.15)=0.60$
5	0.05	$5(0.05)=0.25$
Sum	1.00	$\sum_x x \cdot P(X=x) = 2.55$

Applying the formula of the mean, we get

$$\sum_x x \cdot P(X=x) = 2.55$$

B Standard Deviation for a Discrete Random Variable

The standard deviation of a discrete random variable denoted by σ; is a numerical measure for the spread of the random variable probability distribution.

This numerical measure describes the spread or variability of the random variable using the "average" or "expected value" of the squared deviations of the X-values from their mean μ.

The basic formula to compute the standard deviation for a discrete random variable is

$$\sigma = \sqrt{\sum_x (x-\mu)^2 \cdot P(X=x)}$$

Another formula can be used to calculate the standard deviation of a discrete random variable.

The standard deviation of a discrete random variable X measures the spread of its probability distribution and computed as

$$\sigma = \sqrt{\left[\sum_x x^2 \cdot P(X=x)\right] - \mu^2}$$

A higher value for σ of a discrete random variable indicates that X can assume values over a larger range about the mean. While, a smaller value for σ indicates that most of the values that X can assume are clustered closely about the mean.

The variance (σ^2)of a random variable is obtained by squaring its standard deviation.

Remark

The value of the standard deviation is never negative.

Definition 5.3.2

The standard deviation σ for a random variable X is equal to the square root of its variance.

The variance of a discrete random variable X is the expected value of $(X-\mu)^2$, therefore the formula that can be used to calculate the variance of a discrete random variable is

$$\sigma^2 = \mathrm{E}(X-\mu)^2 = \sum_x (x-\mu)^2 \cdot P(X=x) = \sum_x x^2 \cdot P(X=x) - \mu^2$$

Note that the variance and standard deviation of a discrete random variable are numerical measures describe the spread or variability of the random variable using the "average" or "expected value" of the squared deviations of X-values from their mean μ.

The following examples illustrates how to use the formulas to compute the mean, variance, and standard deviation of a discrete random variable

Example 2

A daily demand for the laptop, is as shown in the table

x	0	1	2	3	4	5
$P(X=x)$	0.1	0.4	0.2	0.15	0.1	0.05

Find the mean, variance, and the standard deviation of X.

Solution: To simplify the calculations, we may construct the following table

x	$\mathrm{P}(X=x)$	$xP(X=x)$	$(X-\mu)^2$	$(X-\mu)^2\,P(X=x)$
0	0.1	0	3.61	0.361
1	0.4	0.4	0.81	0.324
2	0.2	0.4	0.01	0.002
3	0.15	0.45	1.21	0.1815
4	0.1	0.4	4.41	0.441
5	0.05	0.25	9.61	0.4805
Total	1	$\mu = 1.9$		$^2 = 1.79$

Using the table above and applying the formulas of the mean, variance and the standard deviation, we get

$$\sum_x x \cdot P(X = x) = 1.9 \text{ and } \quad ^2 = 1.79.$$

Since the standard deviation can be obtained by taking the square root of its variance, then $\sigma = \sqrt{\sigma^2} = \sqrt{1.79} = 1.34$.

Example 3

In a lottery conducted to benefit the local fire company, 8000 tickets are to be sold \$5 each. The prize is automobile with price \$12000. If you purchase two tickets, what your expected gain.

Solution: Your gain X may take one of two values. You will either lose 10\$ (your"gain" will be -10\$), or win 11990\$, with probabilities, respectively.

$$\frac{7998}{8000} \text{ and } \frac{2}{8000}$$

The probability distribution for the gain is shown in the table

x	-\$10	\$11990
$P(X=x)$	7998/8000	2/8000

The expected gain will be

$$\mu = \mathrm{E}(X) = \sum_x x \cdot P(X = x)$$

$$= -\,10\left(\frac{7998}{8000}\right) + (11990)\left(\frac{2}{8000}\right) = -\ \$7$$

Example 4

Find the mean and the standard deviation of the spots that appear when a die is tossed.

Solution: Let X be denoted to the number of spots, the probability distribution for X is shown in the following table

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

Applying the formula of the mean, then

$$\mu = \mathrm{E}(X) = \sum_{x} x \cdot P(X = x)$$

$$= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)$$

$$= \frac{21}{6} = 3.5$$

To calculate the variance, we apply the following formula;

$$\sigma^2 = \sum_{x} x^2 \cdot P(X = x) - \mu^2$$

$$= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) - (3.5)^2$$

$$= 2.9$$

Hence the standard deviation of X is

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.9} = 1.7$$

Example 5

A box contains 5 balls. Two are numbered 3, one is numbered 4, and two are numbered 5. The balls are mixed and one is selected at random and its number recorded. Then it is replaced, if the experiment is repeated many times, find the standard deviation of the numbers on the balls.

Solution: To find the standard deviation of the numbers on the balls, the following table may be constructed

x	$P(X=x)$	$xP(X=x)$	$x^2P(X=x)$
3	2/5	1.2	3.6
4	1/5	0.8	3.2
5	2/5	2	10
Sum	**1**	$\mu = 4$	**16.8**

To find the standard deviation, first we will find the variance of the numbers on the balls. Applying the formula of the variance, we get

$$\sigma^2 = \sum_x x^2 \cdot P(X = x) - \mu^2$$

$$= 16.8 - (4)^2 = 0.8$$

Taking the square root of the variance, we reach to the standard deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.8} = 0.894$$

Exercises 5.3

1 Let X represents the number of times a customer visits a grocery store in a 1-week period. Assume this is the probability distribution of X:

x	0	1	2	3
$P(X=x)$	0.1	0.4	0.4	0.1

Find each of the following

a The expected value of X

b The standard deviation of X

2 A random variable can assume five values: 0,1,2,3,4 . A portion of the probability distribution is shown here

x	0	1	2	3	4
$P(X=x)$	0.1	0.3	0.3	?	0.1

a Find $P(X=3)$

b Calculate the mean, variance, and the standard deviation for X.

c What is the probability that X is greater than 2.

d What is the probability that X is 3 or less.

3 A random variable has this probability distribution

x	0	1	2	3	4	5
$P(X=x)$	0.1	0.3	0.4	0.1	k	0.05

a Find the value of k, that is $P(X=3)$.

b Find μ, ^{2}, and for X

4 Let X represent the number observed on the throw of a single balanced die.

a Find the probability distribution for X.

b Find μ, ^{2}, and for X

5 A piece of electronic equipment contains six computer chips, two of which are defective. Three chips are selected at random, removed from the piece of equipment, and inspected. Let X represent the number of defective observed, where $X=0,1$ or 2 .Find the probability distribution for X and find for it.

6 The probability that a player will get 5 to ten questions right on a trivial quiz are shown below

x	5	6	7	8	9	10
$P(X=x)$	0.05	0.2	0.4	0.1	0.15	0.1

Find μ, ^{2}, and for X.

7 The probability that a cellular phone company kiosk sells X number of new phones construct per day is shown below

x	4	5	6	8	10
$P(X=x)$	0.28	0.3	0.1	0.17	0.15

Find μ, ^{2}, and for X.

8 A person pays 2$to play a certain game by rolling a single die once. If a 1 or a 2 comes up, the person wins nothing. If, however, the player rolls a 3,4,5 or 6, he wins the difference between the number rolled and 2$. Find the expectation for this game.

9 A civic group sells 1000 rattle tickets to raise 2500$ for its namesake charity. First prize is 1000$, second prize is 300$, and the third prize is 200$. How much should the group charge for each ticket.

10 A recent study of 200 nurses found that of 125 female nurses, 56 had bachelor's degree; and of 75 male nurses, 34 had bachelor's degree. If a nurse is selected at random, find the probability that the nurse is

a A female nurse with a bachelor's degree.

b A male nurse.

c A male nurse with a bachelor's degree.

11 At a convention there are 7 mathematics instructors, 5 computer science instructors, 3 statistics instructors, and 4 science instructors. If an instructor is selected. Find the probability of getting a science or a math instructor.

12 The random variable X is the number of golf balls ordered by customers at a pro shop. Its probability distribution is given in the table.

x	3	6	9	12	15
$P(X=x)$	0.14	0.33	0.36	0.07	0.10

Then any of the following values (A, B, C or D) equal to $E(X)$?

A) 5.31 B) 9 C) 9.54 D) 7.98

5.4 Application to the Random Variable

In this section we will discuss two important distributions as applications to random variables. First part of this section, discusses the binomial distribution as an application to discrete random variable, and the second part discusses the normal distribution as an application to continuous random variable.

The Bernoulli Experiment

A Bernoulli experiment is an experiment satisfying the following conditions:
Each trail has only two outcomes, success with probability ($1>p>0$) or failure probability $q=1-p$.

Examples:

1) A coin tossing experiment is a Bernoulli experiment.

2) The question about the quality of industrial product is a Bernoulli experiment.

3) Surgery to someone's heart is a Bernoulli experiment, either that the operation succeeds and then this person lives or that the operation fails, and then that person die.

Remarks:

1) Usually one symbolizes the result, which represents success and failure with 1 and 0 respectively.

2) If X is a Bernoulli random variable then The Bernoulli distribution is given by the following table:

x	0	1	**sum**
$P(X=x)$	$1-p$	p	1

3) The probability of success (p) is called the parameter of Bernoulli distribution.

4) For a Bernoulli experiment, a success does not mean that the corresponding outcome is considered favorable or desirable one. Similarly, a failure does not necessarily refer to an unfavorable or undesirable outcome. Success and failure are the names used to denote the two possible outcomes of a trial. The outcome to which the question refers is usually called a success; the outcome to which it does not refer is called a failure.

A The Binomial Distribution

The binomial probability distribution is a discrete probability distribution and it is one of the most widely used ones. Such distribution used to find the probability that we get x successes in n performances of an independent Bernoulli experiments.

Now, let us define the binomial probability distribution

Definition 5.4.1

A binomial distribution consists of n independent Bernoulli trials with probability of success p on each trial. The probability of x successes in n independent Bernoulli trials is

$$P(X=x) = nCx \cdot p^{x} q^{n-x} = \frac{n!}{x!(n-x)!} p^{x} q^{n-x}$$

for values of x=0,1,2...,n, where q =1 - p.

Remark

The probabilities for a binomial experiment can also be read from the table of binomial probabilities, as shown in Appendix A. That table lists the probabilities of X for n=1 to n=13 and for selected values of P.

Note that, the natural number n and the real number p are called **parameters** of the binomial distribution.

Let us consider the following examples

Example 1

According to a report issued by the secretary of the emergency department at certain hospital said that 70% of the visitors are men over the weekends. Suppose that 15 visitors are randomly selected over a weekend. Find the probability that exactly 12 of them are men.

Solution: We can solve it by using the definition as follows:

$$P(X=12)={}_{15}C_{12}\,(0.7)^{12}\,(0.3)^{15-12} \approx 0.17$$

Example 2

For a binomial random variable X, with the following informations $n = 10$, $p = 0.1$, x=0,1,2...,10. Find

a. $P(X=2)$ b. $P(X \geq 2)$ c. $P(X \leq 2)$

Solution: We have the following n = 10, p = 0.1. Therefore q=1-0.1=0.9. So

a. $P(X=2) = {}_{10}C_{2}\,(0.1)^{2}(0.9)^{8} = 0.1937$

b. $P(X \geq 2) = 1 - P(x < 2) = 1 - [P(X = 0) + P(X = 1)]$

But

$$P(X = 0) = {}_{10}C_0(0.1)^0(0.9)^{10} = 0.3387$$

and

$$P(X = 1) = {}_{10}C_1(0.1)^1(0.9)^9 = 0.1129$$

Therefore

$$P(X \geq 2) = 1 - [0.3387 + 0.1129] = 0.5484$$

c. $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= 0.3387 + 0.1129 + 0.1937 = 0.6453$$

Example 3

Over a long period of time it has been observed that a given rifleman can hit a target on a single trial with probability equal to 0.8. Suppose he fires 4 shots at the target. Find the following probabilities

a. He will hit the target exactly two times.

b. He will hit the target at least once.

Solution: We have the following $n = 4$, $p = 0.8$. Therefore q=1-0.8=0.2. So

a. $P(X = 2) = {}_4C_2(0.8)^2(0.2)^2 = 0.1536$

b. $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$

$$= 1 - {}_4C_0(0.8)^0(0.2)^4 = 0.9984$$

When a discrete random variable has a binomial distribution, the general formulas used to compute its mean and standard deviation is a "very long"procedure. However, it is simpler and more convenient to use the following formulas

The mean and standard deviation of the binomial random Variable is given by the following formulas:
Mean: $\mu = np$. Variance: $\sigma^2 = npq$ and Standard Deviation: $\sigma = \sqrt{npq}$

The following two examples describe the calculation of the mean and standard deviation of a binomial distribution.

Example 4

A fair coin is tossed 4 times, and we will interest at the number of heads that appears in this experiment. Find the mean, variance, and the standard deviation of the number of heads that will be obtained.

Solution: If X is a random variable describes the appearance of heads then X has a binomial distribution with parameters:

$$n = 4, \; p = 0.5$$

Using these informations we get

$$\mu = np = 4(0.5) = 2, \; \sigma^2 = npq = 4(0.5)(0.5) = 1$$

and

$$\sigma = \sqrt{npq} = \sqrt{4(0.5)(0.5)} = 1$$

Example 5

A fair die is rolled 54 times. If X is a random variable describes the appearance of number 4. Then find the mean, variance, and the standard deviation of X.

Solution: We note that the random variable X has a binomial distribution with parameters:

$$n = 54, \; p = 1/6, \; q = 1 - 1/6 = 5/6$$

Using these informations we get

$$\mu = np = 54(1/6) = 9, \; \sigma^2 = npq = 54(1/6)(5/6) = 7.5$$

and

$$\sigma = \sqrt{npq} = \sqrt{7.5} = 2.73$$

As mentioned at the top of this section, the other application you will learn is the normal random variable and its distribution as an important application on continuous random variable. Moreover, you will learn how to calculate the normal probabilities under certain conditions.

B The Normal Distribution

Many probability distributions that a continuous random variable can possess called the normal distribution which is most commonly used of all probability distributions. In the real life a large number of phenomena are approximately normally distributed. For instance, The continuous random variables representing weights of students or people, scores on an examination, amount of water in a gallon, life age of an item (such as a telephone), have all been observed to have an approximate normal distribution. The distribution of continuous random variables is given by so called density function $f(x)$. The formula of density function of normal distribution is shown below:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

Where e and π are constants given approximately by 2.7183 and 3.1416, respectively; μ and σ $(\sigma > 0)$ are parameters of the distribution, and that represent the mean and standard deviation of it.

The general density function of normal distribution curve is presented in figure 5.1.

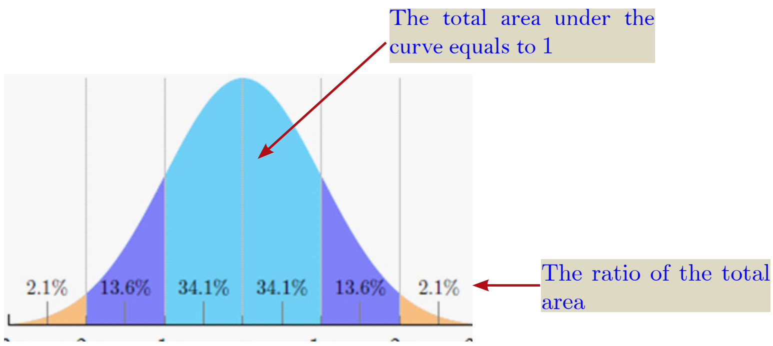


Figure 5.1: The Curve of Density Function of Normal Distribution

According to figure 5.1 we can summarize the most properties of the normal distribution as follows:

1. A normal distribution curve is bell-shaped.

2. The mean, median, and mode are equal and are located at the centre of the distribution.

3. It has only one mode.

4. The curve is continuous and symmetric about the mean.

5. The curve is horizontal asymptotes at y=0.

6. The area under the normal distribution curve is equal to 1.
7. The area under the part of normal curve that lies within 1 standard deviation of the mean is approximately 0.68, within 2 standard deviations about 0.95 and within 3 standard deviations about 0.997.

C The Standard Normal Distribution

A particular case of the normal distribution is called the standard normal distribution. That is the standard normal distribution is normal distribution with zero mean and one unit standard deviation.

Definition 5.4.2

The standard normal distribution is a normal distribution with $\mu = 0$ and $\sigma = 1$.The formula for the density function of standard normal distribution is:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty$$

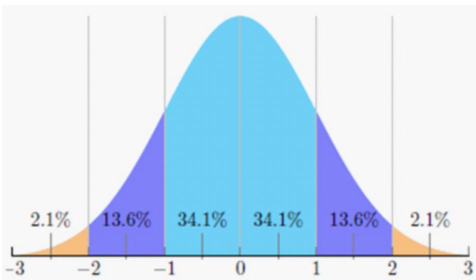


Figure 5.2: The Curve of Density Function of Standard Normal Distribution

Area under the Standard Normal Curve

Some of areas under the standard normal distribution curve are listed in a table called standard normal distribution table, such table 1 in Appendix **A** is listed the areas that lies to the left of a specified values. To illustrate how once can be read the standard normal distribution table, assume we need to find the area to the left of z =1.5. To find this area as follows:

1. Represent the area that needed by graphing a suitable standard normal curve as shown in figure 5.2.

2. Look at row with the value 1.5 in table and for the column with 0.0 0

3. The value inside the table that obtained by intersection the row value and the column value is the area that needed.

Applying the above three steps, then the area to the left of $Z = 1.5$ is 0.9332.

Note that for a random variable Z with standard normal distribution are the following notations may be used:

The area lies between two values a and b denoted by $P(a<Z<b)$

The area lies to the left of the value say a, denoted by $P(Z<a)$

The area lies to the right of the value say b, denoted by $P(Z>b)$

Without loss of generality, the following two cases can be found when calculating the area under the standard normal distribution curve and these cases are covering the most frequencies questions in this issue. The first case is discussed above, but we mention it here in order to summarize all cases that make it easier.

Case 1: Area to the left of $z>0$ we can get it from the table A straight forward.

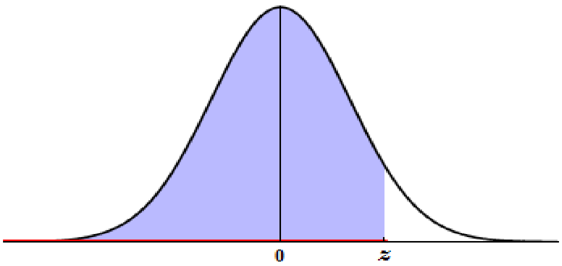


Figure 5.3

Remark

For a continuos random variable X, we have

$$P(X \geq a) = P(X > a)$$
$$P(X \leq a) = P(X < a)$$

Example 6

Find the area to the left of $z = 1.5$ $(z<1.5)$

Solution: Draw the figure 5.4, and put the value 1.5 in the x-axis

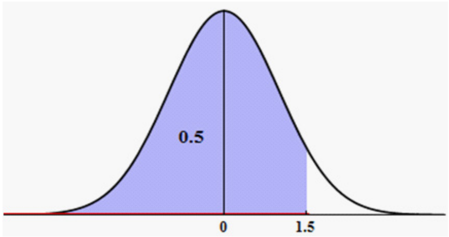


Figure 5.4

According to the standard normal table, the area to the left of z=1.5 is 0.9332.

Case 2: Area to the right of $z>0$. The total area under the standard normal curve is 1, in this case we use the following formula

$$P(Z>z)=1\text{-}P(Z<z)$$

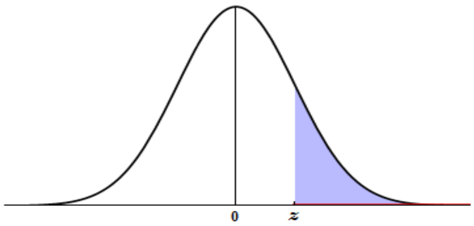


Figure 5.5

Example 7

Find the area to the right of $z = 2$ $(z>2)$

Solution: Draw the figure 5.6, and put the value 2 in the x-axis

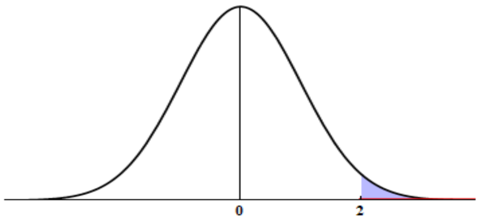


Figure 5.6

Using the formula $P(Z>z)=1\text{-}P(Z<z)$. Then

$$P(Z>2)=1\text{-}P(Z<2)$$
$$=1\text{-}0.9772=0.0228$$

Example 8

Find the area to the left of $z = \text{-}1.5$ $(z < \text{-}1.5)$

Solution: Draw the figure 5.7, and put the value -1.5 in the x-axis

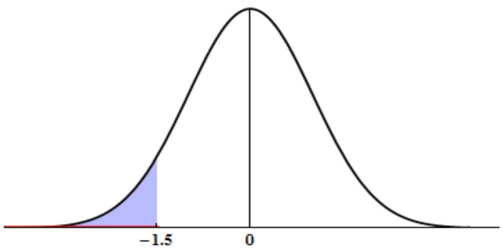


Figure 5.7

Using the table 1 we find $P(Z<\text{-}1.5)= 0.06680$

Example 9

Find the area to the right of z = -0.5 (z >-0.5)

Solution: We have $P(Z>z)$= 1-$P(Z<z)$. So by using table 1 we get

$$P(Z>-0.5)=1- P(Z<-0.5)=1-0.3085=0.6915$$

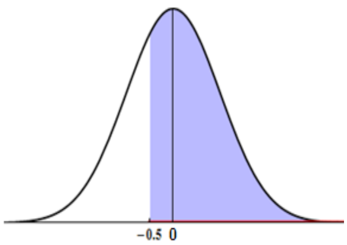


Figure 5.8

Theorem: Let X be a random variable on a space of elementary events S. Then for any two real numbers a and b with $a < b$ we have the following relation:

$$P(a \le X < b) = P(X < b) - P(X < a)$$

If the random variable X is continues then we can write the following relation:

$$P(a < X < b) = P(a \le X < b) = P(X < b) - P(X < a)$$

From the theorem above we can write:

$$P(z_1<Z<z_2)= P(Z<z_2)-P(Z<z_1)$$

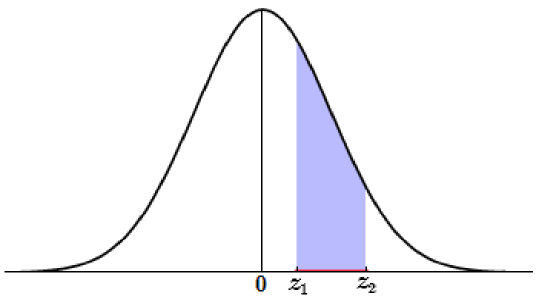


Figure 5.9

Example 10

Find the area between z_1 = 0.5 and z_2= 1.5

Solution: Draw the figure 5.10, and put the values -0.5 and 1.5 in the x-axis

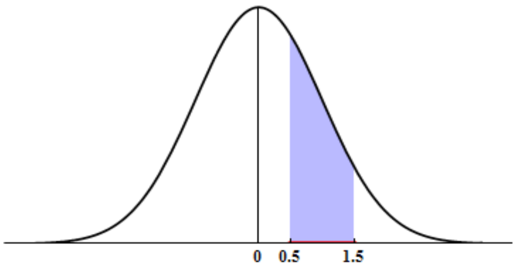


Figure 5.10

We have

$$P(z_1 < Z < z_2) = P(Z < z_2) - P(Z < z_1)$$

Using the table 1 we get

$$P(0.5 < Z < 1.5) = P(Z < 1.5) - P(Z < 0.5)$$
$$= 0.9332 - 0.6915 = 0.2417$$

Example 11

Find the area between z = -1.5 and z = - 0.5

Solution: Draw the figure 5.11, and put the values -1.5 and - 0.5 in the x-axis

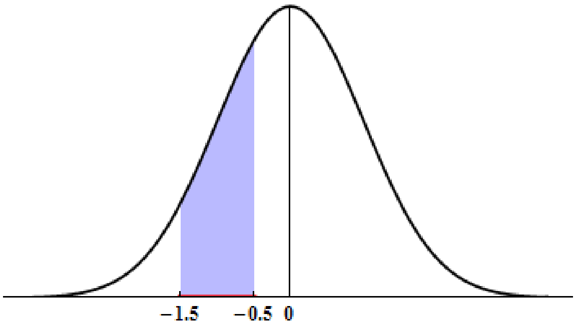


Figure 5.11

We have

$$P(z_1 < Z < z_2) = P(Z < z_2) - P(Z < z_1)$$

So by using the previous theorem and table 1 we get:

$$P(-1.5 < Z < -0.5) = P(Z < -0.5) - P(Z < -1.5)$$
$$= 0.3085 - 0.0668 = 0.2417$$

Example 12

Find the area between z = -0.5 and z =1.5

Solution: Draw the figure 5.12, and put the values -0.5 and 1.5 in the x-axis

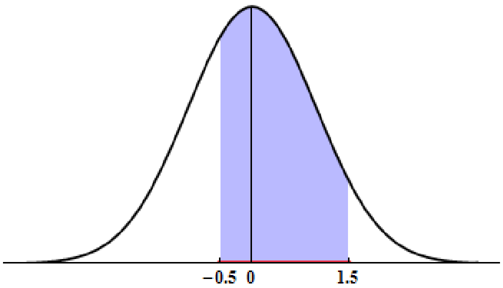


Figure 5.12

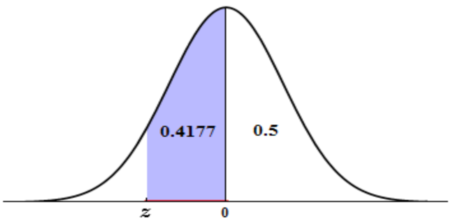
We have $P(z_1<Z<z_2)=P(Z<z_2)-P(Z<z_1)$. So by using the previous theorem and table 1 we get:

$$P(-0.5<Z<1.5)=P(Z< 1.5)-P(Z<-0.5)$$
$$=0.9332-0.3085=0.6247$$

Now we will learn how to find the corresponding value of z when an area under the standard normal distribution curve is known.

Example 13

Find the z value in the figure below



Solution: We can calculate the area to the right of z which is 0.5+0.4177. The table below is a part of the standard normal distribution table. Using this table and since z<0, we get

$$z=-(1.3+0.09)=-1.39$$

z	0.01	0.02	⋯	0.09
⋮				
1.3				0.9177
⋮				

D Standardizing a Normal Distribution

As mentioned earlier that there is a special table to calculate the probabilities or areas under the standard normal distribution curve, whereas real-world problems are normally distributed and each problem has mean and standard deviation different to other, so it is not easy to construct a table for each one, hence no similar table found to calculate an area under the general normal distribution curve. Therefore, a relationship between the standard normal distribution and the general normal distribution is used in order to convert the normal distribution to standard normal distribution and hence the standard normal table is used.

If X is normally distributed with mean μ and standard deviation $\sigma\,(\sigma > 0)$, then corresponding Z of X can be computed by using the formula

$$Z = \frac{X - \mu}{\sigma}$$

The following examples describe how to convert x values to the corresponding standardized z- values and how to find area under a normal distribution curve.

Example 14

Let X be a normally distributed random variable with a mean of 10 and a standard deviation of 2. Find the probability that X lies between 11 and 13.

Solution: The interval from X=11 to X= 13 must be standardized using the formula of Z as follows

For X=11, the corresponding Z value is

$$Z = \frac{X - \mu}{\sigma} = \frac{11 - 10}{2} = 0.5$$

For X= 13, the corresponding Z value is

$$Z = \frac{X - \mu}{\sigma} = \frac{13 - 10}{2} = 1.5$$

The probability that needed is the shaded area in figure 5.13

Using the formula

$$Z = \frac{X - \mu}{\sigma}$$

Then we write

$$P(11<X<13)= P(0.5<Z<1.5)$$
$$= P(Z<1.5)- P(Z<0.5)$$
$$=0.9332-0.6915=0.2417$$

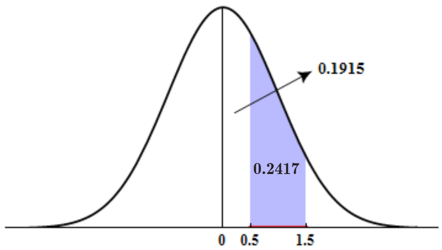


Figure 5.13

Example 15

10000 students take 100-points exam, 6300 of them passed the exam; if the distribution of the scores for all students is normal distribution with mean is equal to 70 and standard deviation 8.

a. Find the number of students gets scores between 60 and 80.

b. What is the score for passing.

Solution:

a. For X=60, the corresponding Z value is

$$Z = \frac{X-\mu}{\sigma} = \frac{60-70}{8} = -1.25$$

For X= 80, the corresponding Z value is

$$Z = \frac{X-\mu}{\sigma} = \frac{80-70}{8} = 1.25$$

Now we can write

$$P(60<X<80)= P(-1.25<Z<1.25)$$
$$=0.8944-0.10560$$
$$=0.7888$$

This mean that the number of the students gets scores between 60 and 80 are 10000(0.7888)=7888.

b. The percentage of the students who have succeeded in the exam is

$$\frac{6300}{10000}.100\% = 63\% = 0.63.$$

To find the passing score, first we will find the z value such that $P(Z>z)=0.63$.

$$P(Z>z)=0.63 \text{ implies } P(Z<z)=1\text{-}0.63=0.37.$$

Using the table of the standard normal distribution we cannot find the value exactly, hence the value 0.3707 is the nearest to 0.37, therefore z =-0.33

Using the formula of Z, then

$$\frac{X-70}{8}=-0.33$$

Solve this equation for X we get

$$X= 70 - 8(0.33) = 67.36 \approx 68$$

This means the passing score is 68

Exercises 5.4

1 Compute the probability of x successes, using the binomial formula

a $n = 5, p = 0.7,\ x = 3.$

b $n = 8, p = 0.6,\ x = 4.$

c $n = 10, p = 0.3,\ x = 7.$

2 Let x be a binomial distribution with n=10, p=0.4. Find the following probabilities

a $P(X=4)$

b $P(X \geq 4)$

c $P(X \leq 4)$

d $P(X<4)$

e Mean and standard deviation of x

3 A student takes a 20-question, True/False exam and guesses on each question. Find the probability of passing if the lowest grade is 15 correct out of 20.

4 A new surgical procedure is said to be successful 80% of times. Suppose the operation is performed five times, and the results are assumed to be independent of one another. What are the probabilities of these events

a All five operators are successful.

b Exactly four are successful.

c Less than two are successful.

d More than 3 are successful

e Find the variance and the standard deviation

5 A study found that 1% of social security recipients are too young to vote, if 800 social security recipients are randomly selected. Find the mean, variance, and the standard deviation of the number of recipients who are too young to vote.

6 If 3% of calculators are defective, find the mean, variance, and standard deviation of a lot of 300 calculators.

7 Thirty-two percent of adult Internet users have purchased products or services online. For a random sample of 200 adult Internet users, find the mean, variance, and standard deviation for the number who have purchased good, or services online.

8 If 13% of businesses have eliminated jobs. If 5 businesses are selected at random, find the probability at least 3 have eliminated jobs during that.

9 Is this a binomial distribution? Explain.

x	0	1	2	3
P(X=x)	0.064	0.288	0.432	0.216

10 A die is rolled 360 times; find the mean, variance, and standard deviation of 1's that will be rolled.

11 Calculate the area under the standard normal curve between these values

a $z = 0$ and $z = 1.6$

b $z = 0$ and $z = 1.83$

c $z = 0$ and $z = 0.95$

d $z = 0$ and $z = 0.90$

12 Calculate the area under the standard normal curve between these values

a $z = -1.4$ and $z = 1.4$

b $z = -2$ and $z = 2$

c $z = -3$ and $z = 3$

13 Find the following probabilities for the standard normal random variable Z

a $P(Z < 2.33)$

b $P(Z < 1.645)$

c $P(Z > 1.96)$

d $P(-2.85 < Z < 2.85)$

14 Find the value of z_0 in each case

a $P(Z > z_0) = 0.025$

b $P(Z < z_0) = 0.9251$

c $P(Z > z_0) = 0.9750$

d $P(Z > z_0) = 0.3594$

e $P(-z_0 < Z < z_0) = 0.8262$

15 A normal random variablec X has mean 10 and standard deviation 2, find the following probabilities

a $P(X > 13.5)$

b $P(X < 8.2)$

c $P(9.4 < X < 10.6)$

16 A normal random variablec X has mean 1.2 and standard deviation 1.5, find the following probabilities

a $P(1<X<1.1)$

b $P(X>1.83)$

c $P(1.35<X<1.5)$

17 A normal random variable has an unknown mean and standard deviation 2 if the probability that exceeds 7.5 is 0.8023, find

18 50 students take 10-points exam, 58 of them passed the exam; if the distribution of the scores for all students is normal distribution with mean is equal to 7 and standard deviation 2.

a. Find the number of students gets scores between 5 and 7.

b. What is the score for passing.

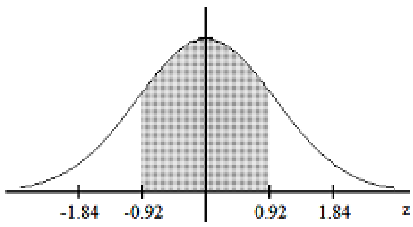
19 For the standard normal curve, the area that lies to the left of 1.13 is

A) 0.8708 B) 0.1292 C) 0.8907 D) 0.8485

20 For the standard normal curve, the area that lies between -1.10 and -0.36 is

A)-0.2237 B)0.4951 C)0.2237 D)0.2239

21 Use the table of areas for standard normal distribution to obtain the shaded area under the standard normal curve.



Any of the following values (A, B, C or D) equal to that area?

A) 0.8212 B) 0.6424 C) 0.3576 D) 0.1788

22 The z-score for which the area under the standard normal curve to its left is 0.96

A) 1.75 B) 1.03 C) 1.82 D) -1.38

23 The random variable X is normally distributed with mean $\mu=15.2$ and standard deviation $\sigma=0.9$. Then any of the following values (A, B, C or D) equal to $P(X>16.1)$?

A) 0.8413 B) 0.1357 C) 0.1550 D) 0.1587

1. Allen, G. Blumann (2010).Elementary Statistics. 7th Ed. London :Mc Graw-Hill International Book Company.

2. Arora, P.N (2008).Comprehensive Statistical Methods. 2nd Ed. India.

3. Barnett, V & Lewis T.(1984).Outliers in Statistical Data. Wiley, New York.

4. Dixon, W.J(1990)BMDP Statistical Software Manual. University of California, Press, Berkery.

5. Downie & R.W. Heath(1983):Basic Statistical Methods. 5th Ed. Harper&Row,New York.

6. Eyank, Nancy(1995).Basic Statistical Methods. 1st Ed. Dar Arcan.

7. Goon, A.M. Gupa & B. Dasgupta (1983)Fundamental of Statistics, 6th Ed. Calcutta, India.

8. Hamburg, M(1977)Statistical Analysis for Decision Making. 2nd Ed. Harcourt Brace Jovanovich, New York.

9. Hawkins, D.M(1980) Identification of Outliers. Champman and Hall, London.

10. Leff, L. S. (2005) . Precalculus The Easy Way. Barron's Educational Series.

11. Hogg, R & Allen Craig (1978)Introduction to Mathematical Statistics. 4th Ed. Collier Macmillan Publishing Company, New York.

12. Lindgren, B.W &G.W(1959)Introduction to Probability and Statistics.1st Ed. Macmillan Publishing Company, New York.

13. Lipschutz, S (1980)Probability. 2nd Ed. Schaum"s Outline Series.1st Ed. Macmillan Publishing Company, New York.

14. Little, R.A and RUBIN D.B (1987) Statistical Analysis with Missing Data. Wiley, New York.

15. Mc Ghee, J. W(1985)Introductory Statistics .West Publish CO. New York.

16. Mansfield, Edwin (1987) Statistic for Business Economics. 7th Ed. PWS Publishers, USA.

17. Moser, C.A & G. Kalton (1989)Survey Methods in Social investigation. 2nd Ed. Gower, England.

18. Prem, S. Mann(2011)Introductory Statistics(International Student Version). 7th Ed. John Wiley&Sons.

19. Robert, J. Beaver & Barbara M.Beaver &William Mendenhall (1999).Introduction to Probability and Statistics. 10th Ed.

20. Salvatori, D (1982) Statistics and Economics. Schaum's Outline Series. Mc Graw-Hill Book Company, New York.

21. Scheaffer, R & W. Mendenhal (1979) Elementary Survey Sampling. 2nd Ed. Duxbury Press, North Scituate, Massachusetts.

22. Sharp, V (1979) Statistics for Social Sciences. Little Brown Company, Toronto, Canada.

23. Spiegel, M (1961) Statistics. Shaum's Outline Series, Mc Graw-Hill Company, New York.

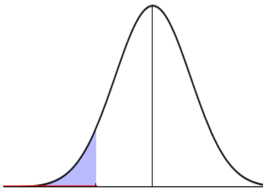
24. Stevenson, W (1978) Business Statistics. Harper & Row Publisher, New York.

25. Weiss Neil (1989) Elementary Statistic. Addison-Wesley Publishing Company, Reading, Massachusetts.

26. Wonnacott T.H & R.W. Wonnacott (1984) Introductory Statistics for Business. 3rd Ed. Tohn Wiley & Sons. New York.

27. Yameni, Taro (1967) Statistics, An Introductory Analysis. 2nd Ed. Harper & Kow Publisher, New York. Seber G.A.F(1984)Multivariate Observations. Wiley, New York.

Table I. The Standard Normal Distribution Table

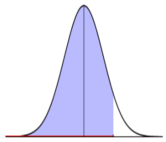


Note: Each table entry is the probability $P(Z < a)$ for a given a, where a is a two decimals number read from the left hand column (for the first decimal), and from the upper horizontal row (for the second decimal).

Example: To find $P(Z < 1.27)$ go down the left column to 1.2, then enter in that row to meet the column under .07. The entry is .8980. So $P(Z < 1.27) = .8980$.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00190	0.00180	0.00180	0.00170	0.00160	0.00160	0.00150	0.00150	0.00140	0.00140
-2.8	0.00260	0.00250	0.00240	0.00230	0.00230	0.00220	0.00210	0.00210	0.00200	0.00190
-2.7	0.00350	0.00340	0.00330	0.00320	0.00310	0.00300	0.00290	0.00280	0.00270	0.00260
-2.6	0.00470	0.00450	0.00440	0.00430	0.00410	0.00400	0.00390	0.00380	0.00370	0.00360
-2.5	0.00620	0.00600	0.00590	0.00570	0.00550	0.00540	0.00520	0.00510	0.00490	0.00480
-2.4	0.00820	0.00800	0.00780	0.00750	0.00730	0.00710	0.00690	0.00680	0.00660	0.00640
-2.3	0.01070	0.01040	0.01020	0.00990	0.00960	0.00940	0.00910	0.00890	0.00870	0.00840
-2.2	0.01390	0.01360	0.01320	0.01290	0.01250	0.01220	0.01190	0.01160	0.01130	0.01100
-2.1	0.01790	0.01740	0.0170	0.01660	0.01620	0.01580	0.01540	0.01500	0.01460	0.01430
-2.0	0.02280	0.02220	0.02170	0.02120	0.02070	0.02020	0.01970	0.01920	0.01880	0.01830
-1.9	0.02870	0.02810	0.02740	0.02680	0.02620	0.02560	0.02500	0.02440	0.02390	0.02330
-1.8	0.03590	0.03510	0.03440	0.03360	0.03290	0.03220	0.03140	0.03070	0.03010	0.02940
-1.7	0.04460	0.04360	0.04270	0.04180	0.04090	0.04010	0.03920	0.03840	0.03750	0.03670
-1.6	0.05480	0.05370	0.05260	0.05160	0.05050	0.04950	0.04850	0.04750	0.04650	0.04550
-1.5	0.06680	0.06550	0.06430	0.0630	0.06180	0.06060	0.05940	0.05820	0.05710	0.05590
-1.4	0.08080	0.07930	0.07780	0.07640	0.07490	0.07350	0.07210	0.07080	0.06940	0.06810
-1.3	0.09680	0.09510	0.09340	0.09180	0.09010	0.08850	0.08690	0.08530	0.08380	0.08230
-1.2	0.11510	0.11310	0.11120	0.10930	0.10750	0.10560	0.10380	0.10200	0.10030	0.09850
-1.1	0.13570	0.13350	0.13140	0.12920	0.12710	0.12510	0.12300	0.12100	0.11900	0.11700
-1.0	0.15870	0.15620	0.15390	0.15150	0.14920	0.14690	0.14460	0.14230	0.14010	0.13790
-0.9	0.18410	0.18140	0.17880	0.17620	0.17360	0.17110	0.16850	0.16600	0.16350	0.16110
-0.8	0.21190	0.20900	0.20610	0.20330	0.20050	0.19770	0.19490	0.19220	0.18940	0.18670
-0.7	0.24200	0.23890	0.23580	0.23270	0.22960	0.22660	0.22360	0.22060	0.21770	0.21480
-0.6	0.27430	0.27090	0.26760	0.26430	0.26110	0.25780	0.25460	0.25140	0.24830	0.24510
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.28100	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.35200	0.3483
-0.2	0.4207	0.4168	0.4129	0.409	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.492	0.488	0.484	0.4801	0.4761	0.4721	0.4681	0.4641

APPENDIX A STATISTICAL TABLES



0 a

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500	0.5040	0.5080	0.5120	0.516	0.5199	0.5239	0.52790	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.56750	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.60640	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.64430	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.68080	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.71570	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.74860	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.77940	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.80780	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.83400	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.85770	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.87900	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.89800	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.91470	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.92920	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.94180	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.95250	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.96160	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.96930	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.97560	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.98080	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.98500	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.98840	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.99110	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.99320	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.99490	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.99620	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.99720	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.99790	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.99850	0.9986	0.9986
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999
3.1	0.99903	0.99906	0.9991	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.9994	0.99942	0.99944	0.99946	0.99948	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.9996	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.9997	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976

Table II. Comulative Binomial Probabilities

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	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	0.8100	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

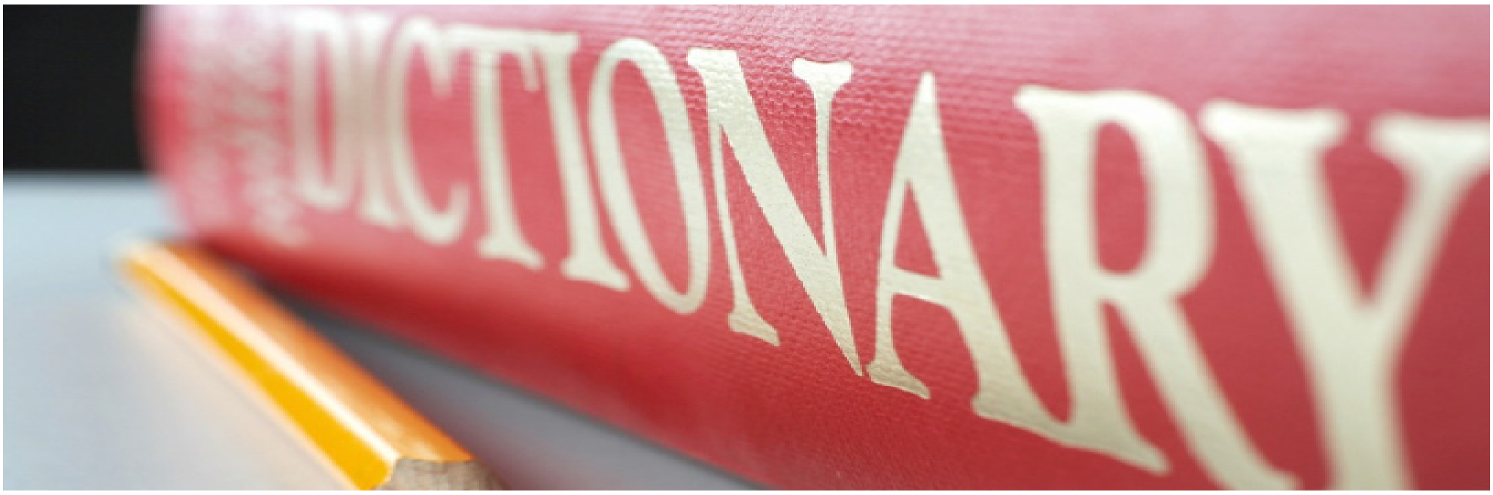
x	p = 0.55	p = 0.60	p = 0.65	p = 0.70	p = 0.75	p = 0.80	p = 0.85	p = 0.90	p = 0.91
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0041	0.0013	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0203	0.0078	0.0025	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

STATISTICAL TABLES

APPENDIX A

APPENDIX A STATISTICAL TABLES


DICTIONARY


integrate
negative
negative multiply
probability
subtraction
infinity

DICTIONARY

A	
Applying	يطبّق
Area	مساحة
Average	معدل
B	
Bar Chart	عرض شرائطي مستطيلات عمودية- أفقية
Basic Concepts	مفاهيم أساسية
Bimodal	ثنائي المنوال
Binomial	حداني او ذو حدان
Boundaries	الحدود الفعلية
Box Plot	مخطط صندوقي
C	
Calculate	أحسب
Categorical	فئوي - تصنيفي
Certain Event	الحادث الأكيد
Class	فئة
Class Width	سعة (طول) الفئة
Cluster Sample	عينة عنقودية
Coefficient of Variation	معامل الاختلاف (أو التغيّر)
Coin	قطعة نقدية معدنية
Collect	جمع
Combination Rule	قاعدة التوافيق
Committee	لجنة
Complementary Events	حوادث متتامة
Compound Event	حوادث مركبة
Conditional	شرطي
Construct	انشيئ
Continuous	مستمر - متصل
Continuous Random Variable	متغير عشوائي مستمر - أو متصل
Convenience Sample	عينة ميسرة
Cumulative	تراكمي - أو تجميعي
Cumulative Frequency	التكرار التراكمي - أو التجميعي
Curve	منحنى
Cumulative Frequency Curve	منحنى التكرار التراكمي - أو التجميعي

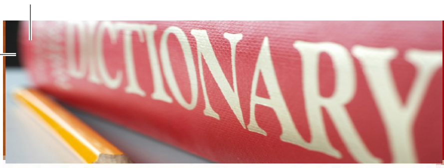
D	
Data	بيانات
Data Recorded	بيانات مسجلة
Data Set	مجموعة بيانات
Data Value Or Datum	قيمة بيان
Dependent	متعلق بـــ - مرتبط بـ
Dependent Variable	متغير تابع
Descriptive Statistics	الإحصاء الوصفي
Die - Dice	حجر نرد - أحجار نرد
Discrete	متقطّع - منفصل
Discrete Random Variable	متغير عشوائي متقطّع - أو منفصل
Discrete Variables	متغيرات متقطّعة - أو منفصلة
Distinguishing	تمييز
Distribution	توزيع
Draw	ارسم
Drawn From	سحب من
E	
Elementary Event	حادث ابتدائي
Equally Likely Events	حوادث لها الإحتمال نفسة
Even Number	عدد زوجي
Events	حوادث
Expected Value	القيمة المتوقعة
Experiment	تجربة
Experiment is Conducted	نفذت التجربة
Experimental Study	دراسة تجريبية
Explanatory Variable	متغير تفسيري
F	
Factorial Notation	صيغة المضروب
Frequency	التكرار
Frequency Distribution	توزيع تكراري
Fundamental Counting Rules	قواعد العد الأساسية
G	
Given That	علماً أن
Graphical	بياني
Graphing	رسم بياني

Greater Than	أكبر من
Greater Than Or Equal	أكبر من أو يساوي
Grouped Data	بيانات مبوبة - أو مجمَّعة
H	
Head	صورة - يقصد فيها صورة الرأس أو الكتابة على قطعة النقود المعدنية
Histogram	المدرج التكراري
I	
Impossible Event	حادث مستحيل
Independent Events	حوادث مستقلة
Independent Variable	متغير مستقل
Inferential Statistics	الإحصاء الاستدلالي - أو الاستقرائي
Interpreting	تفسير
Interquartile Range	المدى الربيعي
Interval Level of Measurement	المستوى الفتراتي للقياس
L	
Largest Value	القيمة الكبرى - أو أكبر قيمة
Lower Boundary	الحد الادنى للفئة العملية
Lower Limit	الحد الأدنى للفئة
Lower True Limit	الحد الأدنى الفعلي للفئة
M	
Mean	المتوسط - أو الوسط الحسابي
Mean Deviation	الإنحراف المتوسط
Mean for Grouped Data	المتوسط لبيانات المبوبة
Mean for Ungrouped Data	المتوسط لبيانات غير مبوبة
Measure Angle	مقياس الزاوية
Measurement Scales	مقاييس القياس
Measures of Central Tendency	مقاييس النزعة المركزية
Measures of Position	مقاييس الموضع
Measures of Variation	مقاييس التشتت
Median	الوسيط
Midpoint	نقطة المركز ويقصد بها مركز الفئة
Midrange	المدى المتوسط
Mode	المنوال

Multimodal	متعدد المنوال
Multiplication Rule	قاعدة الضرب
Mutually Exclusive Events	حوادث متنافية مثنى مثنى
N	
Nominal Level of Measurement	المستوى الاسمي للقياس
Normal Distribution	التوزيع الطبيعي
Numerical Descriptive Measures	المقاييس الوصفية العددية
O	
Occurrence	تَحَقُّق- ظهور
Odd Number	عدد فردي
Ogive	منحنى التكرار المتجمع الصاعد
Ordinal Level of Measurement	المستوى الترتيبي للقياس
Organizing	تنظيم
Outcome	نتائج - ويقصد بها مخرجات التجربة
Outlier Value	قيمة متطرفة
P	
Percentage	النسبة المئوية
Percentile Rank	الرتبة المئينية
Percentiles	المئينات
Permutation Rule	قاعدة التباديل
Pie Chart	الدائرة البيانية
Polygons	المضلّع التكراري
Population	مجتمع إحصائي
Presentation	عرض - أو تمثيل
Probability	علم الاحتمالات
Probabilities	احتمالات
Probability space	فضاء احتمالي
Probability Distribution	توزيع احتمالي
Processed	معالجة
Q	
Qualitative Raw Data.	بيانات خام نوعية
Qualitative Variables	متغيرات نوعية
Quantitative Raw Data.	بيانات خام كمية

Quantitative Variables	متغيرات كمية
Quartiles	ربيعيات
Quasi-Experimental Study	دراسة شبه تجريبية
R	
Random	عشوائي
Random Sample	عينة عشوائية
Random Variable	متغير عشوائي
Range	المدى
Ranked	مرتب - ترتيب
Ratio Level of Measurement	المستوى النسبي للقياس
Raw Data	بيانات خام
Relative	نسبي
Relative Frequency	التكرار النسبي
Replacement	إرجاع - إحلال
Respectively	على التوالي
Respondent	مستجيب
Response	استجابة
Rolling	إلقاء - أو درحرجة - أو رمي
Rolling a Die	إلقاء حجر نرد
S	
Sample	عينة
Set of Elementary Events	مجموعة الحوادث الابتدائية
Sequence	متتالية
Simple Event	حادث بسيط
Smaller Than	أصغر من
Smaller Than or Equal	أصغر من أو يساوي
Smallest Value	القيمة الصغرى - أو أصغر قيمة
Standard Deviation	الانحراف المعياري
Standard Normal Distribution	التوزيع الطبيعي المعياري
Standard Score	علامة معيارية
Statistical Techniques	أساليب إحصائية
Statistics	علم الإحصاء
Stem-And-Leaf	الساق والورقة
Stratified Sample	عينة طبقية
Sum	مجموع
Systematic Sample	عينة منتظمة - أو رتيبة

T	
Tail	شعار - نقش
Tally	التعداد - أو عصا الحساب
Theorem	نظرية
Tossing	قذف
Tossing a Coin	قذف قطعة نقود معدنية
Tree Diagrams	الشجرة البيانية - في الاحتمالات
U	
Ungrouped Data	بيانات غير مبوبة
Unimodal	أحادي المنوال
Upper Boundary	الحد الأعلى للفئة العملية
Upper Limit	الحد الأعلى للفئة
Upper True Limit	الحد الأعلى الفعلي
V	
Variable	متغير
Variance	تباين
Value	قيمة
W	
Weighted Mean	المتوسط الموزون



DICTIONARY