# UNIT 9

# STATISTICS AND PROBABILITY

## Unit Outcomes

By the end of this unit, you will be able to

- **Collect and represent simple statistical data using different methods such as histograms.**
- **Use facts and basic principles of probability.**
- **Develop the concept of probability via experimentation and hypothetical events.**

## Unit Contents


| | | |
|---|---|---|
| | • range | • Measure of central tendency |
| | • median | • measure of dispersion |
| | • sample | • classification of mode |
| • outcomes | • raw data | • descriptive statistics |
| • presentation | • mode | • secondary data |
| • event | • Sample space | • Standard deviation |
| • population | • equally likely | • statistical data |
| • frequency | • tabulation | • frequency distribution |
| • histogram | • variance | • primary data |
| • interpretation | • probability | • variable |
| • analysis | • average | • Arithmetic mean |

## INTRODUCTION

### Activity 9.1

a. Classify the students in your class by their age. What is the average age of the class?

b. Classify the length of time each student in your class takes to get to school. Identify the length of time that is most frequent.

c. Classify time spent to help their parents by each student in your class.

Everybody collects, interprets and uses information; much of it is in numerical or statistical forms in day-to-day life. It is a common practice that people receive large quantities of information everyday through conversations, televisions, computers, the radios, newspapers, posters, notices and instructions. It is just because there is so much information available that people need to be able to absorb, select and reject it. In everyday life, in business and industry, certain statistical information is necessary and it is independent to know where to find it how to collect it. As a consequence, everybody has to compare prices and quality before making any decision about what goods to buy. As employees of any firm, people want to compare their salaries and working conditions, promotion opportunities and so on. In time the firms on their part want to control costs and expand their profits. One of the main functions of statistics is to provide information which will help on making decisions. Statistics provides the type of information by providing a description of the present, a profile of the past and an estimate of the future.

The following are some of the objectives of collecting statistical information.

1. To describe the methods of collecting primary statistical information.
2. To consider the status involved in carrying out a survey.
3. To analyze the observation and interpreting it.
4. To define and describe sampling.
5. To analyze the basis of sampling.
6. To describe a variety of sampling methods.

Statistical investigation is a comprehensive activity and requires systematic collection of data about some group of people or objects, describing and organizing the data, analyzing the data with the help of different statistical methods, summarizing the analysis and using these results for making judgments, decisions and predictions.

## 9.1 Statistical data

**Collection of numerical data.** It is the mathematical science that deals with the gathering, evaluation, and interpretation of numerical facts using the concept of probability of a population from examination of a random pattern.

## 9.1.1 Collection and tabulation of statistical data

**Activity 9.2**

1. Split the class into four groups. Let Group A find the last year's average result of the national exam of all subjects in the school from the school office's records. Let group B collect information about covid-19 treated in your nearest health center, hospital, or health post. Let group C measure the weight of each student in your class. Let group D collect information about the money that each student has in his or her pocket and consider its distribution based on age and sex.

   Answer the following questions using the information gathered by each group.

   a. How many students appeared for the exam?
   b. How many students scored an average mark in the regional exam?
   c. What was the score obtained by most of the students?
   d. Which groups of people are seriously infected by covid-19 diseases?
   e. What is the average weight of the class?
   f. Whose average weight is greater? Males or females?
   g. How many of the students have more than fifty Birr?

There are many definitions of the term statistics given by different scholars. However, for the purpose of this unit, we will confine ourselves to the following:

**Definition 9.1**

Statistics is a science of collecting, organizing, presenting, analyzing and interpreting data so that one can make a generalization.


Figure 9.1: Steps in statistics

## Data collection

In Statistics, data collection is a process of gathering information from all the relevant sources to find a solution to the research problem. It helps to evaluate the outcome of the problem. The data collection methods allow a person to conclude an answer to the relevant question.

## Types of data

### Qualitative data and Quantitative data

Data can be classified as either qualitative or quantitative. However, statistics deals mainly with quantitative data.

Qualitative data is a means for exploring and understanding the meaning individuals or groups ascribe based on some characters whose values are not numbers, such as their color, sex, religion or the town/city they live in.

Quantitative data is a means for testing objective theories by examining the relationships among numerical variables such as scores in exams, height, weight, age or wealth. These variables can be measured, typically on instruments, so that the numbered data can be analyzed using statistical procedures.

### Continuous data and discrete data

Continuous data is information that could be meaningfully divided into finer levels. It can be measured on a scale or continuum and can have almost any numeric value. For example, you can measure your height at very precise scales — meters, centimeters, millimeters and so on.

You can record continuous data at so many different measurements – width, temperature, time, and so on. This is where the key difference with discrete data lies.

Discrete data is a count that usually involves integers. For example, the number of children in a school is discrete data. You can count whole individuals. Some discrete data can be non-integer such as shoes sizes that come in halves.

There are many methods used to collect or obtain data for statistical analysis.

## Here are some data collection methods or instruments:

- Primary data
- Secondary data
- Interviews
- Direct observations
- Questionnaires

### Sources of collecting data

**Primary Data** is a data that has been generated by the researcher himself/herself, surveys, interviews, experiments, specially designed for understanding and solving the research problem at hand.

**Secondary Data** is data used after it is generated by large government institutions, healthcare facilities, etc. as part of organizational record keeping. The data is then extracted from more varied data files.

**Interview:** The main purpose of an interview as a tool of data collection is to gather data extensively and intensively. Exchanging ideas and experiences, eliciting of information pertaining to a very wide range of data in which the interviewee may wish to rehearse his past, define his present and canvass his future possibilities. Interviews can provide in-depth information pertaining to participants' experiences and viewpoints of a particular topic. Generally, interview has two types: a) In-depth and b) focus group.

**Direct observations:** A technique that involves systematically selecting, watching, listening, reading, touching, and recording behavior and characteristics of living things, objects, or phenomena.

Data collection surveys collect information from a targeted group of people about their opinions, behavior, or knowledge. Common types of example **surveys** are written **questionnaires**, face-to-face or telephone interviews, focus groups, and electronic (e-mail or website) surveys.

## Organization of data

The process **of condensing data** and presenting it in a compact form, by putting data into a statistical table, is called **tabulation of Statistical Data**. It is a systematic presentation of numerical data in rows or/and columns according to certain characteristics. It expresses the data in a concise and attractive form which can be easily understood and used to compare numerical figures.

## Presentation of data

The main purpose of data presentation is to facilitate statistical analysis. This can be done by illustrating the data using graphs and diagrams like bar graphs, histograms, piecharts, pictograms, frequency polygons, etc.

## Analysis of data

In order to meet the desired purpose of investigation, data has to be analyzed. The primary objective of analyzing data is to grasp the tendency and characteristics of data from the representative values (mean, median, and mode) and from how it is distributed (dispersion), through the previous steps: organizing and presenting data.

### Exercise 9.1

Classify whether the following data are quantitative or qualitative.

a. Religion
b. Amount of rain fall in a year
c. Number of academic months in a year
d. Monthly income of a person
e. Sex

## Interpretation of data

Based on analyzed data, conclusions have to be drawn. This step usually involves decision making about a large collection of objects (the population) based on information gathered from a small collection of similar objects (the sample).

Some examples that show a government to take action on a certain activity.

**Example 1:** Information gathered about the spread of covid-19 disease in changing trends in health centers and the community.

**Example 2:** Statistics can be used to study the existing disease called kidney failure. It assists in determining the effectiveness of new medication and the importance of counseling.

**Example 3:** Demographic data about religion distribution in Ethiopia and the rate of religious growth, etc., all help policymakers in determining future needs such as providing places for mosques and churches.

**Example 4:** Statistical data collected on telecommunication customers provides feedback that can help to reform an appropriate service.

**Example 5:** Recording annual dropout rates of students in schools that help Ministry of Education of Ethiopia to reform an appropriate policy.

### Population and sampling

Population in statistics refers to all over collection of individuals, objects or measurements that have a common characteristic.

**Census**, an enumeration of people, houses, firms, or other important items in a country or region at a particular time. The term usually refers to a population census. However, many countries take censuses of housing, manufacturing, and agriculture.

Censuses, being expensive, are taken only at infrequent intervals: every 10 years in many countries, every 5 years or at irregular intervals in other countries.

**Sampling frame is the actual set of units from which a sample has been drawn**: in the case of a simple random sample, all units from the sampling frame have an equal chance to be drawn and to occur in the sample.

Reaching access to the whole group (or population) is usually unmanageable, expensive and sometimes destructive. Therefore, instead of examining the whole group, a researcher examines a small part of the group, called a sample.

The following table shows scores of students in mathematics

| Table 9.1: Scores of students in mathematics | |
|---|---|
| Name of student $x$ | Score ( out of 100) $S(x)$ |
| Abdul-aziz | 84 |
| Abebe | 79 |
| Chala | 81 |
| Dendir | 75 |
| Eyasu | 83 |
| Kedija | 77 |
| Meron | 80 |
| Nardos | 72 |
| Obong | 74 |
| Yonas | 82 |

The students are members of the population and their score, S, is the scores in mathematics.

| Table 9.2: Average price of major cereal crop items in Ethiopia | |
|---|---|
| Name of food items | Price in Birr per quintal $P(x)$ |
| Teff | 4500 |
| Barely | 4800 |
| Wheat | 3800 |
| Maize | 3000 |
| Sorghum | 2500 |
| Millets | 4050 |
| Rice | 3500 |

The major cereal crops are members of the population and $p$ is their price.

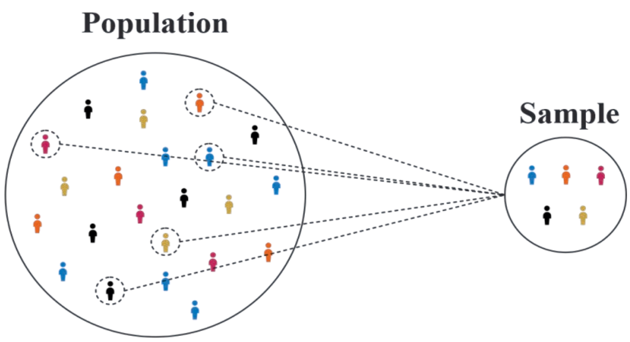



Figure 9.2: sampling

### Example

What are the population and the sample in the following research situations?

**a)** A research randomly selected 1,000 students from different primary schools, wanting to know how much time the Ethiopian students in primary school study per day on average.

**b)** A factory manufactured 100, 000 machines this year. The quality control unit of the factory chooses 250 of them to evaluate the rate of defective products.

**Solution:**

**a)** The total number of primary school students is the population whereas 1,000 students among the population are the sample.

**b)** The population is 100,000 whereas 250 machines are the sample.

## Exercise 9.2

What are the population and sample in the situation below?

a. A researcher wants to know the average income of 5,000 residents of a village. She conducted interviews to 150 people living in the village.

b. A cosmetic company implemented an advertisement that targeted all women in their 20s in Ethiopia. They randomly surveyed 1,000 women in their 20s to examine the effect of the advertisement.

## 9.1.2 Graphical presentations of statistical data

The first step to make a frequency distribution table is to draw two rows on a piece of paper. Then, look through your data set and list all the possible outcomes in the data in the upper rows. Use the bottom row to write for each time that particular outcome occurred in the data.

Graphical representation is another way of analyzing numerical data. A graph is a sort of chart through which statistical data are represented in the form of lines or curves drawn across the coordinated points plotted on its surface. Graphs enable us to study the cause and effect relationship between two variables. Graphs help to measure the extent of change in one variable when other variable changes by a certain amount.

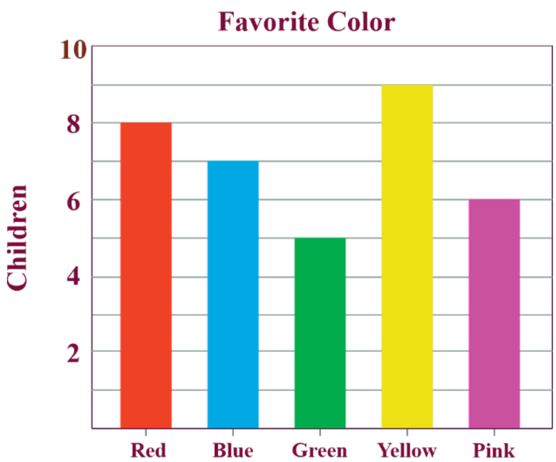



Figure 9.3: Bar chart

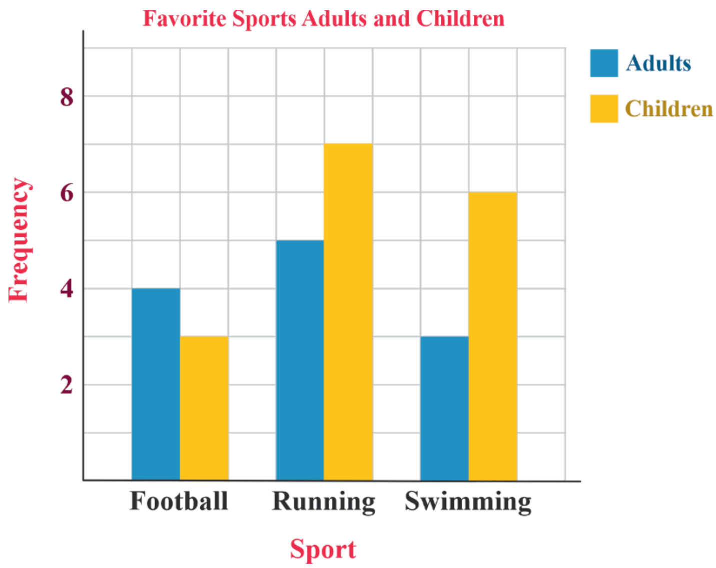



Figure 9.4: Dual Bar

### Pie-charts

Data can be displayed on a pie chart-a circle divided into sectors. The size of the sector is in direct proportion to the frequency of the data. The sector size does not show the actual frequency. The actual frequency can be calculated easily from the size of the sector.

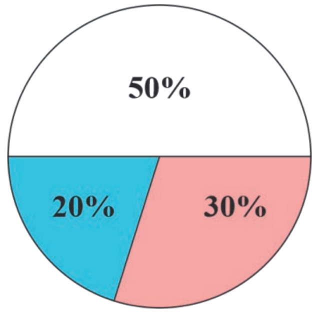



Figure 9.5: Pie-chart

**Example**

In a survey of 120 children, they were asked to choose their favorite color. The total 120 is represented by 100%. It follows that if 120 represent 100%, then

**a.** 50% represents $120\left(\frac{50}{100}\right) = 60$ children whose favorite is white.

**b.** 30% represents $120\left(\frac{30}{100}\right) = 36$ children whose favorite is purple.

**c.** 20% represents $120\left(\frac{20}{100}\right) = 24$ children whose favorite is blue.

## Methods to represent a frequency distribution:

### Distributions and histograms

Data should be organized once it is collected so that it can be manageable. Data that is not organized is called raw data.

**Definition 9.2**

A variable V refers to a characteristic or attribute of an individual or an organization that can be measured or observed and that varies among the people or organization being studied. A variable typically will vary in two or more categories or on a continuum of scores, and it can be measured.

The drawing of tables or graphs enables us to organize raw data and help us manage the data.

### Example 1

Suppose there are 12 people in a village whose weights in kilograms were measured as follows: $55, 62, 49, 67, 55, 62, 62, 49, 67, 62, 55, 49$. These data are raw. Organize the data using a table. What are the variables?

| Weight in kg ( V) | 49 | 55 | 62 | 67 |
|---|---|---|---|---|
| Number of people ( $f$ ) | 3 | 3 | 4 | 2 |

The above table is called the frequency distribution table. The variables are the weight and the number of people is the frequency.

### Example 2

There is an age data of 20 customers who visited a shopping center. $2,4,11,12,18,22,25,28,30,30,31,34,38,39,45,48,49,50,60,64$. Complete the frequency table.

**Solution:**

Let $x$ be the age of the customers. Divide the data into groups of 10 years. Each interval in the frequency distribution table is called class. This type of frequency table is called grouped frequency distribution table in grade 11 Unit 7.

| Class interval | frequency |
|---|---|
| $0 \le x < 10$ | 2 |
| $10 \le x < 20$ | 3 |
| $20 \le x < 30$ | 3 |
| $30 \le x < 40$ | 6 |
| $40 \le x < 50$ | 3 |
| $50 \le x < 60$ | 1 |
| $60 \le x < 70$ | 2 |

## Exercise 9.3

Complete the frequency tables with the data below.

a) 1, 1, 1, 2, 3, 3, 4, 4, 4, 4, 5, 5, 6, 6, 6

| Value | frequency |
|---|---|
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |

b) 2,3,4,6,8,9,10,11,13,13,15,17

| Class interval ($x$) | frequency |
|---|---|
| $0 \leq x < 5$ | |
| $5 \leq x < 10$ | |
| $10 \leq x < 15$ | |
| $15 \leq x < 20$ | |

## Definition 9.3

**A** histogram is a graphical representation of a frequency distribution in which the variable $(V)$ is plotted on the $x$-axis and the frequency $(f)$ is plotted on the $y$-axis.

## Steps in drawing a histogram

1. Draw a horizontal line. This will be where we denote our classes.
2. Place evenly spaced marks along this line that correspond to the classes.
3. Label the marks so that the scale is clear and give a name to the $x$- axis.
4. Draw a vertical line just to the left of the lowest class.
5. Choose a scale for the vertical axis that will accommodate the class with the highest frequency.
6. Label the marks so that the scale is clear and give a name to the $y$- axis.
7. Construct bars for each class. The height of each bar should correspond to the frequency of the class at the base of the bar. We can also use relative frequencies for the heights of our bars.

### Example

Plot the following mathematics scores ($s$) by a histogram

| Class interval | frequency |
|---|---|
| $25 \le s < 30$ | 2 |
| $30 \le s < 35$ | 2 |
| $35 \le s < 40$ | 5 |
| $40 \le s < 45$ | 10 |
| $45 \le s < 50$ | 6 |
| $50 \le s < 55$ | 2 |
| $55 \le s < 60$ | 3 |

**Solution:**

In this graph, we shall take class intervals on the $x$- axis and frequencies in the $y$-axis. Before plotting the graph, we have to convert the class into their exact limits. Histogram plotted from the data

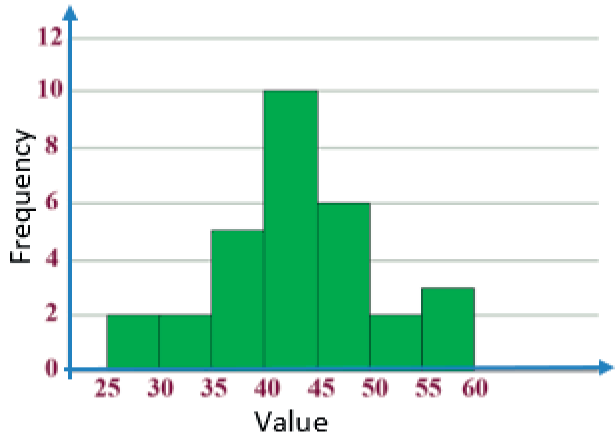



Figure 9.6: Histogram

## Activity 9.3

A group of students collected the data on the height in cm of one type of plant. Here is the data of 30 pieces of the plant.

61, 63, 64, 66, 68, 69, 71, 71.5, 72, 72.5, 73, 73.5, 74, 74.5, 76, 76.2, 76.5, 77, 77.5,

Construct a frequency distribution table and a histogram for these data.

## Exercise 9.4

1. Construct a histogram for the following frequency distribution table that describes the frequencies of weights of 25 students in a class.

| Weight(w) in kgs | Frequency (number of students) |
|---|---|
| $45 \leq w < 50$ | 4 |
| $50 \leq w < 55$ | 10 |
| $55 \leq w < 60$ | 8 |
| $60 \leq w < 65$ | 3 |

2. Complete the frequency distribution table and draw the corresponding histogram for the class interval by 5cm from 145cm. The row data is given as follows: 145, 148, 150, 157,160, 164, 166, 169, 171

| Height(x) | Frequency (number of students) |
|---|---|
| $145 \leq x < 150$ | |
| $150 \leq x < 155$ | |
| $155 \leq x < 160$ | |
| $160 \leq x < 165$ | |
| $165 \leq x < 170$ | |
| $170 \leq x < 175$ | |

## 9.1.3 Measures of central tendency

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the median and the mode.

The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. In the following sub-units, we will look at the mean, mode and median, and learn how to calculate them and under what conditions they are most appropriate to be used.

### Mean

**Activity 9.4**

1. A teacher gave a mid-term exam to 20 students. The full score of this exam is 25. The students scored the following:

   9,9,9,10,11,11,12,12,12,12,13,14,14,14,14,14,18,18,18,21

   a. Calculate the mean score of the students.

   b. How many students score below the mean, exactly the mean and above the mean?

   c. How well the students did the exam?

## Definition 9.4

The Arithmetic mean (or average) is the most popular and well-known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data (see our Types of Variable guide for data types). The mean is equal to the sum of all the values in the data set divided by the number of values in the data set. So, if we have n values in a data set and they have values $x_1, x_2, \dots \dots . x_n$ the mean, usually denoted by $\bar{x}$ (pronounced " $x$ bar"), is:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} .$$

The mean is essentially a model of your data set. It is the value that is most common. You will notice, however, that the mean is not often one of the actual values that you have observed in your data set. However, one of its important properties is that it minimizes error in the prediction of any one value in your data set. That is, it is the value that produces the lowest amount of error from all other values in the data set.

The mean can also be calculated from its frequency distribution. So, if the values $x_1, x_2, \dots \dots . x_n$ occur $f_1, f_2, f_3, \dots \dots f_n$ times, respectively, then the mean ($\bar{x}$ ) is given by

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \cdots \dots x_n f_n}{f_1 + f_2 + f_3 + \cdots \dots + f_n}$$

### Example 1

Seven children have the following number of textbooks: 5, 4, 6, 4, 3, 6 and 7.

Calculate the mean of the textbooks of the children.

**Solution:**

$$\bar{x} = \frac{5+4+6+4+3+6+7}{7} = \frac{35}{7} = 5$$

## Example 2

Calculate the mean test score from the frequency distribution table below that shows the test scores of 10 students.

| Test scores | Frequency |
|---|---|
| 10 | 0 |
| 20 | 4 |
| 30 | 1 |
| 40 | 3 |
| 50 | 2 |

**Solution:**

$$\bar{x} = \frac{10(0)+20(4)+30(1)+40(3)+50(2)}{10} = \frac{330}{10} = 33$$

## Example 3

Ten people have the following number of pencils: 6, 5, 6, 5, 3, 6, 6, 10, 10 and 3.

Calculate the mean of the pencils of the people.

**Solution:**

| No. of pencils ($x$) | Frequency ($f$) |
|---|---|
| 3 | 2 |
| 5 | 2 |
| 6 | 4 |
| 10 | 2 |

$$\bar{x} = \frac{3(2)+5(2)+6(4)+10(2)}{10} = \frac{60}{10} = 6$$

## Exercise 9.5

a) Calculate the mean of the weights in kg of 6 people below

45, 50, 55, 60, 70, 80

b) Calculate the mean height of the people shown in the frequency distribution below.

| Height (cm) | Frequency |
|---|---|
| 160 | 3 |
| 170 | 2 |
| 175 | 1 |

## Properties of mean

### Activity 9.5

Six students have the following amount of money: Birr 55, 45, 60, 40, 30 and 70.

a. Calculate the mean of money in the group.
b. Find the sum of the differences of each value from the mean.
c. If one gives Birr 5 to each student, calculate the new mean.
d. If each student gives Birr 5 to somebody from what they have, calculate the new mean.
e. If each of their money is multiplied by 5, what is the new mean?
f. Discuss the answers you got in a, b, c, d and e

**Property 1:** The sum of the deviations from the mean is zero.

If $\bar{x}$ is the arithmetic mean of $n$ observations $x_1,\ x_2, \dots \dots . x_n$, then

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) \dots \dots . . + (x_n - \bar{x}) = 0$$

Proof: $\bar{x} = \frac{x_1 + x_2, \ldots\ldots. + x_n}{n}$

$$x_1 + x_2 + \cdots \ldots. + x_n = n\bar{x} \ldots \ldots \ldots \ldots \ldots \ldots \ldots *$$

$(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) \ldots \ldots.. + (x_n - \bar{x}) = x_1 + x_2, \ldots \ldots. + x_n - n\bar{x} = n\bar{x} - n\bar{x} = 0, \ldots \ldots \ldots \ldots.. Using *$

Hence, $(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) \ldots \ldots.. + (x_n - \bar{x}) = 0$

## Property 2:

The mean of $n$ observations $x_1, x_2, \ldots \ldots. x_n$ is $\bar{x}$. If each observation is increased by $p$, the mean of the new observations is $(\bar{x} + p)$.

Proof: $\bar{x} = \frac{x_1 + x_2 +, \ldots\ldots. + x_n}{n}$

$$x_1 + x_2 + \cdots \ldots. + x_n = n\bar{x}$$

The mean of $(x_1 + p), (x_2 + p), (x_3 + p) \ldots \ldots., (x_n + p)$

$$= \frac{(x_1+p) + (x_2+p) + (x_3+p) + \cdots \ldots. +, (x_n+p)}{n}$$

$$= \frac{(x_1 + x_2 + \cdots \ldots. + x_n) + np}{n} = \frac{x_1 + x_2 + \cdots \ldots. + x_n}{n} + \frac{np}{n} = \bar{x} + p$$

Hence, the mean of the new observations is $\bar{x} + p$.

## Property 3:

The mean of n observations $x_1, x_2, \ldots \ldots. x_n$ is $\bar{x}$. If each observation is decreased by $p$, the mean of the new observations is $\bar{x} - p$.

## Property 4:

The mean of n observations $x_1, x_2, \ldots \ldots. x_n$ is $\bar{x}$. If each observation is multiplied by a nonzero number $p$, then the mean of the new observations is $p\bar{x}$.

$$\bar{x} = \frac{\mathrm{x}_1 + \mathrm{x}_2, \ldots\ldots. + \mathrm{x}_\mathrm{n}}{\mathrm{n}}$$

$$x_1 + x_2 + \cdots \ldots. + x_n = n\bar{x}$$

The mean of $x_1p,\ x_2p, \dots\dots x_np\ = \frac{(x_1p)+(x_2p)+,\dots\dots\dots+(x_np)}{n}$

$$= \frac{(x_1+\ x_2+\cdots\dots+x_n)p}{n} = p\bar{x}.$$

## Property 5:

The mean of n observations $x_1,\ x_2, \dots\dots x_n$ is $\bar{x}$. If each observation is divided by a nonzero number $p$, then the mean of the new observations is $\frac{\bar{x}}{p}$.

## Exercise 9.6

1. If $\bar{x} = 4$, find the new mean ín case
   a) each value increases by 1.
   b) each value decreases by 2.
   c) each value is multiplied by 3.
   d) each value is divided by 2
2. The weekly mean number of customers who use a bus stop is 100 per day. If 10 more customers use it every day, what is the weekly mean number of the customers?
3. A car factory manufactures 1,000 cars every day on average. One day a new machine is installed, and they can now produce twice as many cars every day. What is the mean number of cars that they produce every day?

## Median

### Definition 9.5

The median is the middle score for a set of data that has been arranged in order of magnitude. The median is less affected by extreme values and skewed data.

### Example 1

The following are the scores of 11 people. Find the median.

65, 55, 89, 56, 35, 14, 56, 55, 87, 45, 92

**Solution:**

We first need to rearrange that data into order of magnitude (smallest first):

14, 35, 45, 55, 55, **56**, 56, 65, 87, 89, 92

Our median mark is the middle mark - in this case, **56** (highlighted in bold). It is the middle mark because there are 5 scores before it and 5 scores after it. This works fine when you have an odd number of scores, but what happens when you have an even number of scores? What if you had only 10 scores? Well, you simply have to take the middle two scores and average the result. So, if we look at the example below:

65, 55, 89, 56, 35, 14, 56, 55, 87, 45

We again rearrange that data into order of magnitude (smallest first):

14, 35, 45, 55, **55**, **56**, 56, 65, 87, 89

Now, we have to take the $5^{th}$ and $6^{th}$ scores in our data set and average them to get a median of 55.5.

$$\text{Median} = \frac{55+56}{2} = 55.5$$

**Properties of the median**

1. It is not affected by extreme values.
2. It is unique for a given data set.

**Example 2**

Find the median of the numbers $5, 3, 99, x, 4$ where $x$ is greater than 10 less than 99.

**Solution:**

Arranging in increasing order as $3, 4, 5, x, 99$, hence, 5 is the median and is not affected by extreme value, 99 and is unique.

## Exercise 9.7

a) Find the median of the data below.

5, 7, 3, 10, 1, 5, 9

b) Calculate the mean and median and compare the values of the two for the data below.

15, 10, 2, 6, 7, 20, 3, 18, 100

## Mode

### Definition 9.6

The **mode** is the most frequent value in a set of data.

### Example 1

Find the mode of the following data.

5, 10, 20, 20, 30, 35, 35, 35, 40, 50, 70, 70

**Solution:**

The most frequent value is 35. Hence, it is the mode.

### Example 2

A teacher gave a mid-term exam to 20 students. The full score of this exam is 25%. The students scored the following:

18, 9, 11, **14**, 11, **12**, 13, 13, **12**, 9, 10,
**14**, **12**, **14**, **12**, **14**, 18, 18 , 21, 9

**Solution:**

The most frequent value in the data are **14** and **12.** Therefore the mode are 14 and 12. On a histogram, mode represents the highest bar in a bar chart or histogram. You can, therefore, sometimes consider the mode as being the most popular option. An example of a mode is presented below:

### Example 3

Find the mode of the following data. 1, 3, 6, 7, 9, 8, 10, 17, 18

## Solution:

Here, there is no most frequent value. No value repeats itself. Hence, there is no mode.

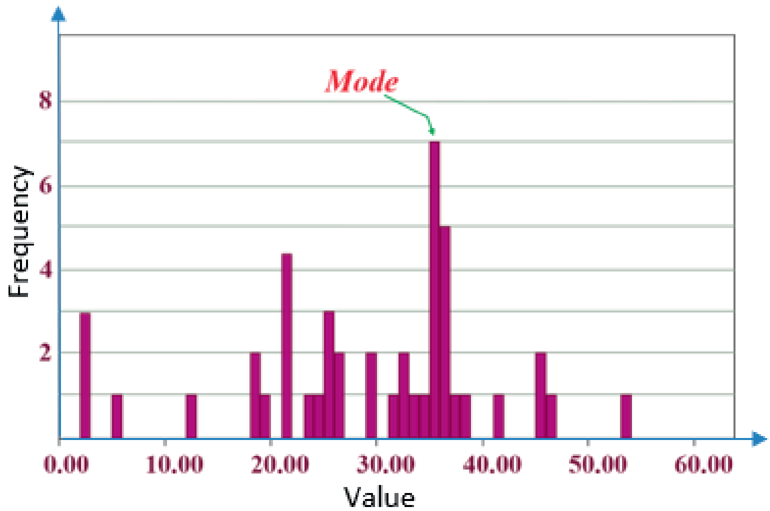



Figure 9.7: Mid exam scores of grade 9 students

Normally, the mode is used for categorical data where we wish to know which is the most common category, as illustrated below:

However, one of the problems with the mode is that it is not unique, so it leaves us with problems when we have two or more values that share the highest frequency, such as below.

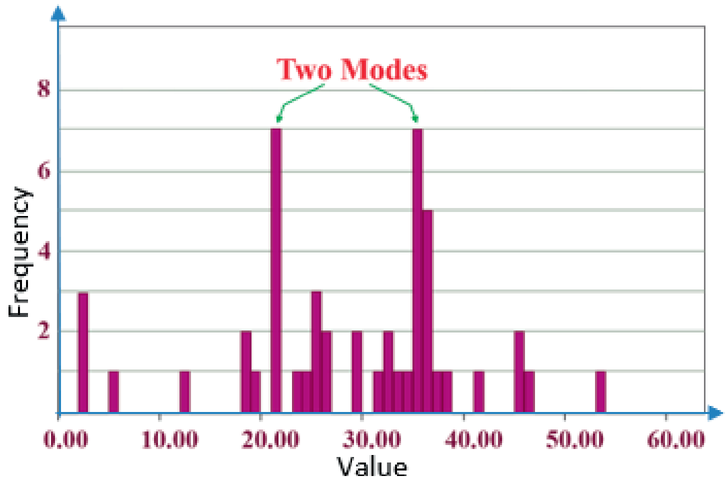



Figure 9.8: The final –exam score of grade 9 students

## Properties of the mode

1. The mode is not always unique
2. The mode can also be used for qualitative data

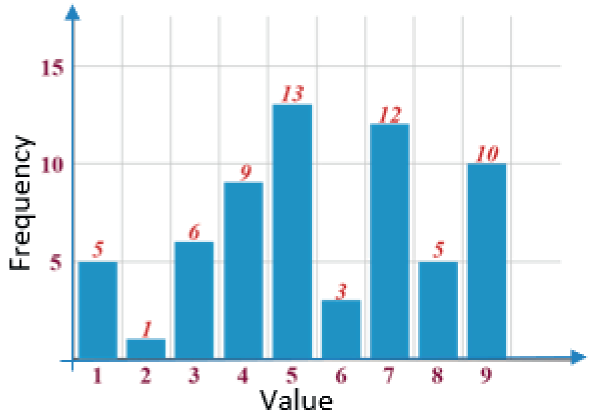
## Exercise 9.8

Find the mode of the following data.

a) 2, 3, 3, 4, 5, 5, 5

b) 13, 10, 16, 15, 12, 14, 13, 16

c) Find the mode from the bar chart on the right




## Exercise 9.9

1. A boy recorded the number of book pages that he read over the last 5 days. 24, 10, 16 ,7, 33
   a. Find the mean number of pages read
   b. Find the median
   c. Find the mode
2. A woman was recording her weight in kg in the last six months as follows: 88, 86, 89, 91, 85, 90
   a. What is the mean weight of the woman in the last six months?
   b. Find the median
3. Is it possible to find the arithmetic mean of qualitative data?
4. Record car accidents in your neighboring city/town to your school together with their teacher.
   a. Find the mean car accident in the city/town

b. Find the median

c. Find the most frequent (highest frequency) in the accident

5. Find the mean, median and mode of the following set of numbers:

$82, 23, 59, 94, 70, 26, 32, 83, 87, 94$, and $32$

6. Answer the following by referring to the frequency distribution given.

a. Find the mean, median and mode

b. How many of the values are greater than or equal to 2?

| $V$ | $-4$ | $-1$ | $0$ | $1$ | $2$ | $3$ |
|---|---|---|---|---|---|---|
| $f$ | 2 | 3 | 4 | 5 | 3 | 3 |

7. If the mean is 5 for the data $2, 3, x, 5, 6, 12$, find the value $x$.

8. If the mean of $a, b, c, d$ is $k$, then what is the mean of $a + b, 2b, c + b, b + d$?

9. If the mean of $y + 2, y + 4, y + 6$ and $y + 10$ is $13$, find the value of $y$.

10. In an examination, the mean of marks scored by a class of 40 students was calculated as 72.5. Later on, it was detected that the marks of one student were wrongly copied as 48 instead of 84. Find the correct mean.

11. The mean monthly salary of 12 employees of a firm is Birr 1450. If one more person joins the firm who gets Birr 1645 per month, what will be the mean monthly salary of 13 employees?

## 9.1.4 Measures of dispersion

When comparing sets of data, it is useful to have a way of measuring the scatter or spread of the data.

**Activity 9.6**

Consider the following four sets of numbers representing the ages of children

| Group | Values | | | | | | | | | | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| I | 5 | 7 | 6 | 7 | 4 | 5 | 4 | 6 | 8 | 5 | |
| II | 2 | 1 | 4 | 3 | 9 | 6 | 3 | 2 | 4 | 3 | |
| III | 6 | 5 | 7 | 5 | 6 | 7 | 5 | 6 | 5 | 7 | |
| IV | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | |

1. Complete the table by finding the sum of each group and by calculating the mean.
2. Are the means equal?
3. Compare the difference between the mean and each observed value in Group $I, II, III$ and $IV$.
   i. Which group's mean is closest to each value?
   ii. Which group has the greatest difference between the mean and each data value?
4. Compare the variation of each group?
   i. Which group shows most variation?
   ii. Which group shows no variation?
   iii. Which group shows slight variation?
5. Find the range for each group.

Dispersion or Variation is the scatter (or spread) of data values from a measure of central tendency. There are several measures of dispersion that can be calculated for a set of data. In this sub-unit, we will consider only three of them, namely, the range, the variance and the standard deviation.

## 1. Range

**Activity 9.7**

Given the raw data as follows: 6, 8, 2, 10, 5, 8, 11, 7. Find the difference between the largest and the smallest.

The simplest and the crudest measure of the dispersion of quantitative data is the range

## Definition 9.7

The difference between the largest and smallest values of a set of numerical data is called the **range R** .i.e. Range = Largest value − Smallest value.

### Example 1

The scores of 10 students in mathematics test is as follows: 6, 7, 5, 8, 3, 8, 9, 5, 4, 5.5. Find the range of the scores of the students.

**Solution:**

Range = Largest value – Smallest value = $9 - 3 = 6$

### Example 2

Find the range of the frequency distribution from the table below:

| $V$ | 1 | 2 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $f$ | 3 | 6 | 8 | 2 | 1 |

**Solution:**

Range = Largest value – Smallest value = $6 - 1 = 5$

## Exercise 9.10

a) The weights of a family member composed of 5 persons are as follows: Find the range. 35,40,85,65,70

b) The table below shows the number of customers who came to a shop in the week. Find the range.

| Day | Mon | Tue | Wed | Thu | Fri | Sat | Sun |
|---|---|---|---|---|---|---|---|
| Number of customer | 30 | 23 | 45 | 17 | 5 | 41 | 54 |

## 2. Variance and standard deviation

The activities below assist you to learn how to find variance and standard deviation.

### Activity 9.8

Consider the following numerical data: $4, 5, 7, 8, 7, 5$

a. Find the mean

b. Find the difference of each value from the mean (deviation)

c. Square each of the deviations

d. Calculate the mean of these squared deviations and its principal square root

### Definition 9.8

**Variance** is the mean of the squared deviations of each value from the arithmetic mean and is denoted by $\sigma^2$

### Definition 9.1

**Standard deviation** is the square root of variance.

**How to calculate standard deviation can be summarized as the following steps:**

1. Calculate the arithmetic mean $\bar{x}$ of the data
2. Find the deviation of each data from the mean $(x - \bar{x})$
3. Calculate the square of each of these deviations, $(x - \bar{x})^2$
4. Find the mean of these squared deviations. This value is the variance ($\sigma^2$). Let n be the number of observations or cases
5. $\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$
6. Take the principal square root of the variance, ($\sigma^2$), i.e.

   Standard deviation $(\sigma) = \sqrt{\text{variance}}$,

   $$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}.$$

### Example 1

Calculate the variance, $\sigma^2$ and standard deviation, σ if the number of televisions sold in each day of a week is $13, 8, 4, 9, 7, 12, 10$.

**Solution:**

$$\bar{x} = \frac{13+8+4+9+7+12+10}{7} = \frac{63}{7} = 9$$

| $x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ |
|---|---|---|
| **13** | **4** | **16** |
| **8** | **-1** | **1** |
| **4** | **-5** | **25** |
| **9** | **0** | **0** |
| **7** | **-2** | **4** |
| **12** | **3** | **9** |
| **10** | **1** | **1** |

Variance, $\sigma^2 = \frac{16+1+25+0+4+9+1}{7} = \frac{56}{7} = 8$

Standard deviation, $\sigma = \sqrt{\sigma^2} = \sqrt{8} \cong 2.83$

## Exercise 9.11

1. With the data $1, 3, 5, 5, 7, 9$
   a. Calculate the mean
   b. Calculate the variance
   c. Calculate the standard deviation
2. With the data $5, 2, 3, 3, 7$ and their arithmetic mean, $\bar{x} = 4$
   a. Find the variance
   b. Find the standard deviation

## Example 2

48 students were asked to write the total number of hours per week they spent watching television. With this information, find the standard deviation of hours spent watching television

**Solution:**

| Hours (h) per week | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|
| $f$ | 3 | 6 | 9 | 13 | 8 | 5 | 4 |

$$\bar{x} = \bar{x} = \frac{6{\cdot}3+7{\cdot}6+8{\cdot}9+9{\cdot}13+10{\cdot}8+11{\cdot}5+12{\cdot}4}{48}$$

$$= \frac{432}{48} = 9$$

| $x_i$ | $f_i$ | $x_i f_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $f_i(x_i - \bar{x})^2$ |
|---|---|---|---|---|---|
| 6 | 3 | 18 | -3 | 9 | 27 |
| 7 | 6 | 42 | -2 | 4 | 24 |
| 8 | 9 | 72 | -1 | 1 | 9 |
| 9 | 13 | 117 | 0 | 0 | 0 |
| 10 | 8 | 80 | 1 | 1 | 8 |
| 11 | 5 | 55 | 2 | 4 | 20 |
| 12 | 4 | 48 | 3 | 9 | 36 |
| | $n = 48$ | | | | |

$$\sigma^2 = \frac{27 + 24 + 9 + 0 + 8 + 20 + 36}{48} = \frac{124}{48} = \frac{31}{12}$$

Hence, $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{31}{12}} \cong \sqrt{2.58} \cong 1.6$

## Exercise 9.12

A student asked his 20 classmates how many books they read every month.

| Number of books | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Frequency($f$) | 6 | 9 | 4 | 1 |

Complete the table and find the variance and the standard deviation.

| $x_i$ | $f_i$ | $x_i f_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $f_i(x_i - \bar{x})^2$ |
|---|---|---|---|---|---|
| 1 | 6 | 6 | | | |
| 2 | 9 | 18 | | | |
| 3 | 4 | 12 | | | |
| 4 | 1 | 4 | | | |
| | $n = 20$ | | | | |

## Properties of standard deviation

### Activity 9.9

Consider the following data which shows the amount of rain fall in mm in the last five days. $20, 30, 50, 70, 80$

1. Find the mean.
2. Calculate the variance and standard deviation.
3. In the next five days, if the rainfall increases by 5mm each, i.e. $25, 35, 55, 75, 85$.
   a. Find the mean of rainfall in those five days.
   b. Find the variance and standard deviation of rainfall in those five days.
   c. Compare the answers with those found in 1 and 2 above.
   d. Discuss the comparison you found in c.

The above group work assists you in noticing the following properties:

**1.** Property 1.

If a constant $k$ is added to each value of a data, then the new variance is the same as the old variance. The new standard deviation is also the same as the old.

**Proof:**

Let $x_1,\ x_2, \dots \dots . x_n$ be n observations with mean $\bar{x}$ and variance $\sigma^2$. Adding to each observation, we get $x_1 + k,\ x_2 + k, \ldots \ldots x_n + k$, then the new mean is $\bar{x} + k$.

New variance $= \frac{[(x_1+k)-(\bar{x}+k]^2+[(x_2+k)-(\bar{x}+k]^2+\cdots+[(x_n+k)-(\bar{x}+k]^2}{n}$

$= \frac{[x_1-\bar{x}]^2+[x_2-\bar{x}]^2+\cdots+[x_n-\bar{x}]^2}{n} = \sigma^2$ and the new standard deviation is σ

## Example

Consider a set of data 2, 3, 4, 3

**a.** Calculate the variance and standard deviation.

**b.** Add 3 to each value, and then find the new variance and the new standard deviation.

**Solution:**

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

$$x - \bar{x}: -1, 0, 1, 0 \text{ and } (x - \bar{x})^2: 1, 0, 1, 0$$

**a.** $\sigma^2 = \frac{1+0+1+0}{4} = \frac{2}{4} = 0.5$ & $\sigma = \sqrt{0.5} \cong 0.7071$

**b.** Adding 3: we have the data $5, 6, 7, 6$

Let $\bar{x}_1$ be the original mean and $\bar{x}_2$ be the new mean.

$\bar{x}_2 = \frac{5+6+7+6}{4} = \frac{24}{4} = 6$

Or using Property 2 of mean,

$\bar{x}_2 = \bar{x}_1 + \text{p} = 3 + 3 = 6$ $\quad$ $x - \bar{x}: -1, 0, 1, 0$ and $(x - \bar{x})^2: 1, 0, 1, 0$

The new variance is $\sigma^2 = \frac{1+0+1+0}{4} = \frac{2}{4} = 0.5$ & $\sigma = \sqrt{0.5} \cong 0.7071$

The new variance = the old variance and the new standard deviation = the old standard deviation.

## Exercise 9.13

a) Find the variance and standard deviation of the following data. $3, 9, 2, 7, 4$

b) If you add 2 to each value of a), how will the values of variance and standard deviation change?

c) If you subtract 2 from each value of a), how will the values of variance and standard deviation change?

Property 2.

If each value of data is multiplied by a constant $c$, then

**a.** The new variance is $c^2$ times the old variance

**b.** The new standard deviation is |c| times the old standard deviation.

**Proof:**

Consider $n$ observations $x_1, x_2, \ldots \ldots . x_n$, the variance $\sigma^2$ and the mean $\bar{x}$.

Multiplying each data value by $c$, we have $c\bar{x}$. The new variance

$$= \frac{(cx_1 - c\,\bar{x})^2 + (cx_2 - c\,\bar{x})^2 + \cdots \ldots\ldots + (cx_n - c\,\bar{x})^2}{n}$$

$$= \frac{c^2(x_1 - \bar{x})^2 + c^2(x_2 - \bar{x})^2 + \cdots \ldots \ldots \ldots + c^2(x_n - \bar{x})^2}{n}$$

$= c^2$ times the old variance $= c^2\sigma^2$

Therefore, the new standard deviation $= |c|\sigma$

## Exercise 9.14

a. Find the variance and standard deviation of the following data: $3, 16, 1, 12$

b. Find the new variance and standard deviation after multiplying each value by 2

c. Find the new variance and standard deviation after dividing each value by 7.

## Exercise 9.15

1. Find the mean, median, and mode(s) of the following data $3, 4, 5, 5, 7, 8, 9, 9, 10$

2. Find the range, variance and standard deviation of the following data.

   5, 5,2,5,1, 4,4,3,6,5,3,5.

3. Find the range, variance and standard deviation of the distribution in the table below:

| $V$ | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| $f$ | 3 | 2 | 1 | 8 | 1 | 2 | 3 |

4. If the standard deviation of the data $5, y,\ 5, 5, 5$ *is* 5, what is the value of $y$?
5. If the variance of $a, b, c, d, e$ *is* $m$, then calculate
   a. The variance of $2a, b+a, c+a, d+a$ *and* $e+a$.
   b. The standard deviation of $2a, b+a, c+a, d+a$ *and* $e+a$.
   c. The standard deviation of $a^2, ab,\ ac,\ ad$ *and* $ae$.
6. The daily rainfall for the last five days was $60, 90, 150, 210, 240$. For the next five days, the amount of rainfall was doubled each, i.e., $120, 180, 300, 420,$ *and* $480$.
   a. Calculate the mean, variance and standard deviation of the doubled rain fall.
   b. Compare the above result $(6a)$ with questions number 1 and 2 of activity 9.9.
7. Calculate the variance, $\sigma^2$ and standard deviation, $\sigma$ if the number of exercise books sold in each day of a week is 9, 8, 4, 12, 5, 14, 11.
8. Twenty students were asked to write the total number of hours per week they spent watching Ethiopian television. With this information find the standard deviation of hours spent watching Ethiopian television.

| Hours (h) per week | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| $f$ | 2 | 6 | 5 | 7 |

## 9.2 Probability

### Introduction

In your grade 8 lessons, you came across with the word probability as you often apply it," The probability of passing an exam'', or "there is a high probability that they can win the game", and so on. Here, the concept probability describes guesses of possibilities. Recall what you have discussed in grade 8 on this lesson based on the group work below:

## Activity 9.10

1. Which of the following is different from the others?

   a. Chance b. Interpretation c. Possibilities d. Uncertainty.

2. All the possible outcomes that can occur when a coin is tossed twice are listed on a paper: $HH, HT, TH, TT$

   What is the probability of having at least a head?

Probability is a numerical value that describes the likelihood of the occurrence of an event in an experiment.

## Definition 9.10

An experiment is a trial by which an observation is obtained but whose outcome canot be predicted in advance.

**Experimental probability** is a probability determined by using data collected from a repeated data.

## Definition 9.11

If an experiment has $n$ equally likely outcomes and if $m$ is a particular outcome, then the probability of this outcome occurring is $\frac{m}{n}$

### Example

All the possible outcomes that can occur when a coin is tossed twice are $HH, HT, TH, TT$. What is the probability of having

**a)** Exactly one tail?

**b)** At least one head?

**Solution:**

**a)** Two ( $HT$ & $TH$ ) out of the four possibilities. Hence, probability $= \frac{2}{4} = \frac{1}{2}$

**b)** Three ( $HH, HT$ & $TH$ ) out of the four possibilities. Hence, probability $= \frac{3}{4}$

## Exercise 9.16

1. All the possible outcomes that can occur when a coin is tossed twice are $HH, HT, TH, TT$. What is the probability of having
   a) Exactly one head?
   b) No tail?
2. List all the possible outcomes.
   a) A fair coin is tossed once
   b) A fair coin is tossed three times

## Sample space and event

### Definition 9.13

Sample space for an experiment is the set of all possible outcomes of an experiment.

### Example 1

**a.** List the sample space in tossing one coin.

**b.** Give the sample space in throwing one die.

**c.** What is the sample space in throwing two coins?

**Solution:**

**a.** In tossing one coin, there are only two possibilities: Heads $(H)$ or Tails $(T)$ . Therefore, $S = \{H, T\}$.

**b.** In throwing one die, any one of the numbers $1, 2, 3, 4, 5, 6$ will appear on the upper face of the die. Hence, $S = \{1,2,3,4,5,6\}$.

**c.** The set of possible outcomes is $S = \{HH, HT, TH, TT\}$.

## Definition 9.14

An **event** is a subset of a sample space set or a sample space.

## Activity 9.11

Duguma throws a fair die once. According to this experiment, discuss the following

Figure 9.9: a die

1. Is it possible to guess the number that shows on the upper face of the die? Why?
2. Write the sample space.
3. From the experiment, give an example of an event.
4. Which event (a) or (b) is certain and which is impossible?
   a) The number on the upper face of the die is 8.
   b) The number on the upper face of the die is a natural number.
5. Determine the possibilities of the following events.
   i. The number on the upper face of the die is 6
   ii. The number on the upper face of the die is 0
   iii. The number on the upper face of the die is less than or equal to 4
6. Discuss the following words:
   a. Experiment
   b. Sample space
   c. Event
   d. Impossile event
   e. Certain event

### Example 2

The numbers 4 to 30 are each written on one of 27 identical cards. One card is chosen at random.

**i.** List the elements of all the possible outcomes

**ii.** Write all the possible elements of the following events

**a.** the number is greater than or equal to 4

**b.** the number is less than or equal to 30

**c.** the number is greater than 15

**d.** the number is divisible by 3

**e.** the number is even

**Solution:**

**i.** $S = \{4, 5, 6, \dots \dots .. 30\}$

**ii.** Write all possible elements of the following events:

**a.** $S = \{4, 5, 6, \dots \dots .. 30\}$

**b.** $S = \{4, 5, 6, \dots \dots .. 30\}$

**c.** $S = \{16, 17, 18 \ \dots \dots .. 30\}$

**d.** $S = \{6, \ 9, 12, \dots \dots .. 30\}$

**e.** $S = \{4, 6, \ 8 \dots \dots .. 30\}$

## Exercise 9.17

The numbers 3 to 30 are each written on one of 28 identical cards. One card is chosen at random.

i. List the set of all possible outcomes

ii. Write all the possible elements of the following events

a) the number is greater than or equal to 5

b) the number is less than or equal to 15

## Probability of an event

### Activity 9.12

A coin is tossed three times, six times, nine times and twelve times and the observation is recorded as follows:

| | Number of tosses | | | | Total |
|---|---|---|---|---|---|
| The number of times the coin is tossed | 3 | 6 | 9 | 12 | 30 |
| The number of times the coin shows up heads | 1 | 4 | 4 | 6 | 15 |
| The number of times the coin shows up tails | 2 | 2 | 5 | 6 | 15 |

What proportion of the number of tosses shows Heads? Tails? What is the probability that the outcome is Head? Tail?

If you toss a coin 10 times and get a head 6 times and a tail 4 times, then you would say that in a single toss of a coin, the probability of getting a head is $\frac{6}{10} = 0.6$. Once again, if you toss a coin 1,000 times and get a head 600 times and a tail 400 times, then you would say that in a single toss of a coin, the probability of getting a head is $\frac{600}{1,000} = 0.6$ . We may obtain different probabilities for the same event from various experiments. Repeating the experiment sufficientely large number of times, however, the relative frequency of an outcome will tend to be close to the theoretical probability of that outcome.

### Example 1

Toss a coin 10,000 times and you obtain 5010 heads,

i. If the event was tails, how many times did this event occur?

ii. According to the experiment, what was the probability of tails?

**Solution:**

**i.** The times you get a tail is:
10000 – 5010 = 4990
4990 times

**ii.** 4990/10000 = 0.499
Probability of tails is 0.499

## Exercise 9.18

a. Toss a coin 10,000 times and you obtain 5610 heads.

**i.** If the event was tails, how many times did this event occur?

**ii.** According to the experiment, what was the probability of tails?

b. You toss a die 100 times, and each number appeared as follows:

1. 20 times
2. 16 times
3. 15 times
4. 14 times
5. 18 times
6. 17 times

i) If the event is 1, how many times did the event occur?

ii) According to the experiment, what is the probability of getting 2?

iii) What is the probability of getting 5 or 6?

iv) What is the probability of getting an odd number?

### Example 2

In an experiment of choosing workers at random, a supervisor found the following result after 100 trials.

| Gender of worker | Men | Women | Total |
|---|---|---|---|
| Number of workers | 45 | 55 | 100 |

What is the probability that a randomly selected worker is a man?

**Solution:**

The probability that a randomly selected worker is a man will be the ratio of the number of men to the total number of trials.

The probability of a man selected $= \frac{45}{100}$

## Exercise 9.19

1. In a class, there are 20 boys out of 40 classmates in total. In an experiment of choosing a classmate at random, what is the probability of choosing a boy?
2. There are 12 males and 8 females in a room. If one person is randomly selected, what is the probability that the selected one is a female?
3. A bag contains 5 red balls, 2 blue balls, and 3 black balls. Find the probability of getting a black ball if you pick a ball from the bag at random.
4. Suppose you are planning a research for public and private schools and wants to sample 1 school at random. If the ratio of the number of public schools to that of private schools is 9:1, what is the probability of sampling a private school?

## Tree diagram

When more than two combined events are being considered, two-way tables cannot be used and therefore another method of representing information diagrammatically is needed. Tree diagrams are a good way of doing this. A tree diagram is one way of showing the possible outcomes of a repeated experiment.

### Example 1

In an experiment of tossing three coins,

**a.** What are the possible outcomes?

**b.** How many different possible outcomes are there?

**c.** What is the probability of the coin facing up with

  **i.** Exactly two heads?

ii. Exactly one tail?

iii. Three tails

iv. At least two heads?

**Solution:**

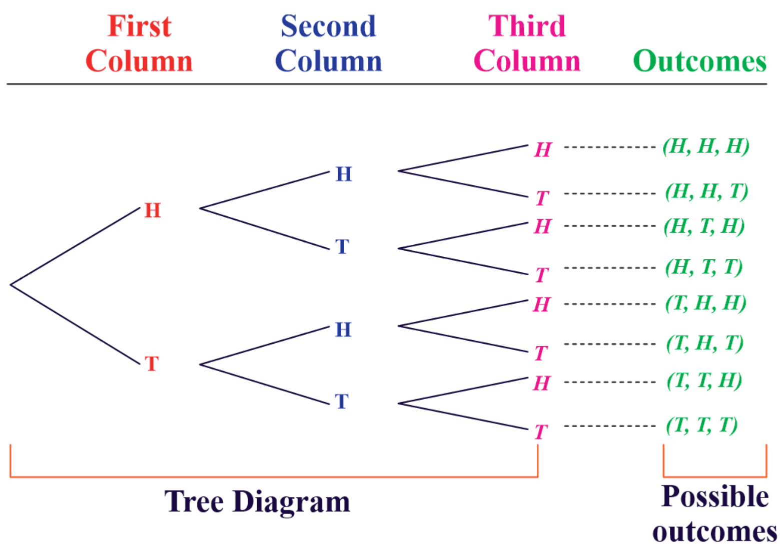



Figure 9.10: A tree diagram in tossing three coins

a. S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}.

b. As can be seen from a) there are 8 possible outcomes.

c. The probability of the coin facing up with is:

i. Exactly two heads $= \frac{3}{8}$.

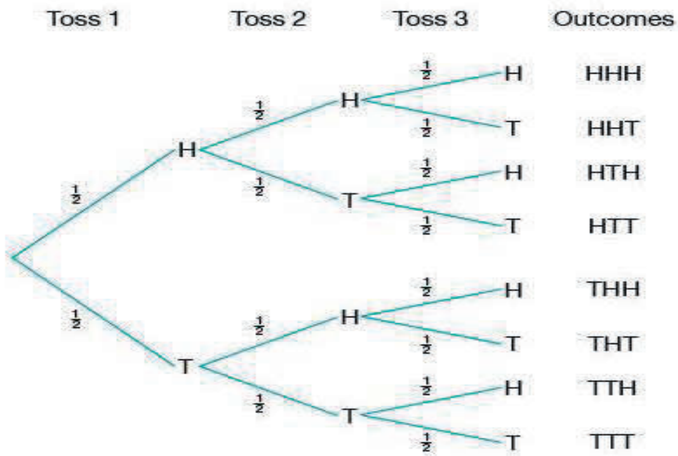
ii. Exactly one tail $= \frac{3}{8}$.

iii. Three tails $= \frac{1}{8}$.

iv. At least two heads$= \frac{4}{8} = \frac{1}{2}$.

## Example 2

If a coin is tossed three times, the tree diagram is as follows. Each of the probabilities is indicated at the side of the branches.




## Exercise 9.20

Calculate the probability of getting the following events when three fair coins are tossed by drawing a tree diagram.

a. Exactly one head
b. Three heads
c. At least two tails

## Theoretical probability

Note that it is not always possible to perform an experiment and calculate probability. In such cases, we should have another method to calculate the probability of an event. We will see a theoretical approach to find such probabilities.

### Definition 9.15

The theoretical probability of an event $E$, written as $P(E)$ is defined as follows:

$$P(E) = \frac{\text{Number of outcomes favourable to the event E}}{\text{Total number of possible outcomes(S)}} = \frac{n(E)}{n(S)}$$

### Example 1

If you throw a die once, what is the probability that an odd number will show on the upper face of a die?

**Solution:**

Let the event of odd number be E

$S = \{1,2,3,4,5,6\}, \ E = \{1,3,5\}$

$$P(E) = \frac{\text{Number of outcomes favourable to the event E}}{\text{Total number of possible outcomes(S)}} = \frac{3}{6} = \frac{1}{2}$$

### Example 2

A coin and a die are thrown together.

**a.** Sketch a tree diagram showing the outcomes of this experiment.

**b.** What is the probability of getting a tail and an odd number?

**c.** What is the probability of getting a head and a number less than or equal to 5?

**Solution:**

**a.**

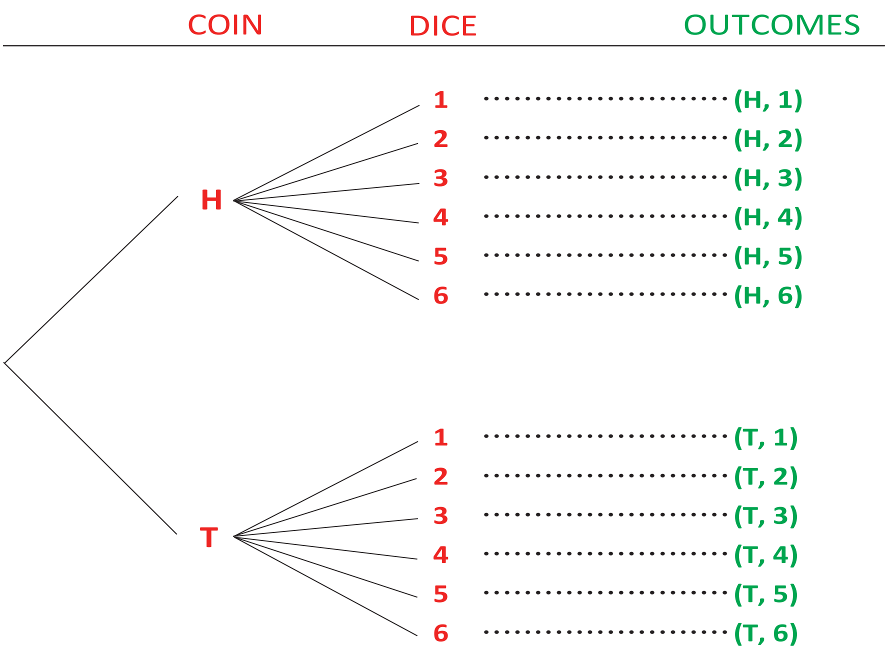



Figure 9.11: A tree diagram in tossing a coin and a die

**b** Probability of getting a tail and an odd number $= \frac{3}{12} = \frac{1}{4}$

**c** Probability of getting a head and a number less than or equal to 5 is $= \frac{5}{12}$

## Exercise 9.21

1. A coin and die are tossed together.

   a) Find the probability of getting a head and even number.

b) Find the probability of getting a number less than 2

2. A bag contains a red, blue and black ball. You run an experiment where you pick a ball from the bag at random and put it back to the bag after confirming which color is picked. You run this experiment twice.
   a) Find the probability of getting a red ball once
   b) Find the probability of getting two balls of the same color
   c) Find the probability of getting two balls of different color

## Exercise 9.22

1. You toss a die once.
   a. What is the sample space?
   b. Write all possible elements if the event is getting less than 4.
2. Find the probability of not getting a tail when you toss two fair coins together.
3. Two dice are simultaneously thrown once. Write all possible elements of the following events.
   a. The sum of the numbers facing up is 8.
   b. The product of the two numbers facing up is 2.
   c. Consecutive numbers facing up
4. Three coins are tossed at the same time. Sketch a tree diagram for the outcomes of this experiment. Write the sample space.
5. There are 30 adults and 20 kids in a room. If you randomly select 1 person from the room, what is the probability of selecting a kid?
6. A bag contains five white balls, four black balls and six red balls. A ball is drawn out of the box at random. What is the probability of drawing the ball for:
   a. Red? b. black? c. white?
7. Two dice are thrown together. What is the probability that the number obtained on one of the dice facing up is a multiple of the number obtained on the other dice facing up?

## Summary

1. Statistics is the science of collecting, organizing, presenting and interpreting data in order to have a conclusion.
2. A population is a complete collection of individuals, objects or measurements that have common characteristics.
3. A subset of a population is called a sample.
4. With **descriptive statistics**, your goal is to describe the data that you find in a sample or is given in a problem.
5. **Primary Data** is a data that has been generated by the researcher himself/herself, surveys, interviews, experiments, specially designed for understanding and solving the research problem at hand.
6. **Secondary Data is** a data used after it is generated by large government institutions, healthcare facilities etc. as part of organizational record keeping. The data is then extracted from more varied data files.
7. If a classification is not based on numbers, then it is called a qualitative classification.
8. If a classification is based on numbers, then it is called a quantitative classification.
9. With **inference statistics**, your goal is to use the data in a sample to draw conclusions about a larger population.
10. A statistical table is a systematic presentation of data in columns and rows.
11. A histogram is a graphical representation of a frequency distribution in which the variable $(V)$ is plotted on the x-axis and the frequency $(f)$ is plotted on the y-axis.
12. A frequency distribution is a distribution showing the number of observations associated with each data value.
13. The **mean** is the arithmetic average of the data.

**14.** If $x_1, x_2, \dots \dots . x_n$ are $n$ observations, then the mean, $\bar{x} = \frac{x_1 + x_2 + \cdots \dots + x_n}{n}$.

**15.** The **median** is the number in the middle of a dataset. When the data has an odd number of counts, the median is the middle number after the data have been ordered. When the data has an even number of counts, the median is the average of the two most central numbers.

**16.** The **mode** is the most often occurring value in the data.

**17.** The outcomes of an experiment are said to be equally likely if, when the experiment is repeated a large number of times, each outcome occurs equally often.

**18.** The sample space for an experiment is the set of all possible outcomes of the experiment.

**19.** If $S$ is the sample space of an experiment and each element of $E$ is equally likely to occur, then the probability of the event $E$ occurring, denoted by $P(E)$, is defined as:

$$P(E) = \frac{\text{number of elements in E}}{\text{number of elements in S}} = \frac{\text{n(E)}}{\text{n(S)}}$$

## Review Exercise

1. What does presenting and tabulating data mean?
2. The amount of money in birr that 20 people have in their pocket are given below: 4,9,2,3,9,1,2,4,2,4,6,6,7,2,5,9,3,5,8,9
   a. Construct a frequency distribution table
   b. What percent of the people have less than Birr 4?
   c. Draw a histogram to represent the data.
3. A bag contains nine balls distinguishable only by their colors; four are blue and five are red. You draw one ball and record their colors.
   a. What is the sample space?
   b. Express the event that the ball drawn is red as a subset of the sample space.
4. A fair coin is tossed, and a fair die is thrown. Write down sample spaces for
   a. tossing of the coin;
   b. throwing of the die;
   c. combination of these experiments.
   d. Let $A$ be the event that a head is tossed, and $B$ be the event that an odd number is thrown. Calculate the probability of the occurrence of A and B by referring to the sample space. Calculate the probability of the occurrence of A or B by referring to the sample space.
5. Refer to the following information to answer the questions that follow.
   Find the **mean, median, mode, range, variance and standard deviation** for the mean temperatures recorded over a five-day period last winter:
   $18, 22, 19, 25, 12.$
6. Find the **mean**, **median, mode, range, variance and standard deviation** of the scores of an exam out of 20% whose distribution is given in the table below.

| $v$ | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|
| $f$ | 6 | 7 | 5 | 7 | 3 | 2 |

**7.** Which of the following is true?

**A.** The mean, mode and median of a data cannot be equal.

**B.** The range of a data cannot be a non-positive number.

**C.** The sum of the deviations of each value of a population from the mean will always be zero.

**8.** A fair die is rolled. What is the probability that a $1, 4, 5$, or $6$ will be on the upper face?

**9.** A spinner is divided into 3 equal parts, with parts labeled $8, 9$, and $10$.

**i.** What is the probability of spinning an 8 on the spinner if you know the arrow landed on an even number

**ii.** What is the probability of spinning 9 on the spinner if you know the arrow landed on an odd number

**10.** A pair of dice is rolled. Find the probability that the product of the numbers on the upper faces is:

**i.** 1

**ii.** less than 6

**iii.** odd

**iv.** greater than or equal to 15

**v.** less than 2

**11.** A card is drawn from a well shuffled pack of 52 cards. What is the probability of getting queen of club or a king of hearts card?

**12.** The points received from the first 7 games of a football club were respectively 1, 3, 1, 3, 3, 1, and 3. What must be the point in the 8$^{th}$ game to have a mean of 2 points?

**A.** 0 **B.** 1 **C.** 3 **D.** No answer

**13.** Refer to the frequency distribution table below to answer question that follows:

| $v$ | 9 | 6 | 5 | 4 | 3 | 2 |
|---|---|---|---|---|---|---|
| $f$ | 6 | 7 | 5 | 7 | 3 | 2 |

Which of the following is true?

**A.** the mean is 5.4

**B.** the mode is 9

**C.** the median is 5.5

**D.** No answer

# Trigonometric table

Corresponding angles (in degree)

| | sin | cos | tan | | |
|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | | 90 |
| 1 | 0.017452 | 0.999848 | 0.017455 | 57.2900 | 89 |
| 2 | 0.03490 | 0.999391 | 0.034921 | 28.63628 | 88 |
| 3 | 0.052336 | 0.99863 | 0.052408 | 19.08115 | 87 |
| 4 | 0.069756 | 0.997564 | 0.069927 | 14.30068 | 86 |
| 5 | 0.087156 | 0.996195 | 0.087489 | 11.43006 | 85 |
| 6 | 0.104528 | 0.994522 | 0.10510 | 9.514373 | 84 |
| 7 | 0.121869 | 0.992546 | 0.122784 | 8.144353 | 83 |
| 8 | 0.139173 | 0.990268 | 0.140541 | 7.115376 | 82 |
| 9 | 0.156434 | 0.987688 | 0.158384 | 6.313757 | 81 |
| 10 | 0.173648 | 0.984808 | 0.176327 | 5.671287 | 80 |
| 11 | 0.190809 | 0.981627 | 0.19438 | 5.144558 | 79 |
| 12 | 0.207912 | 0.978148 | 0.212556 | 4.704634 | 78 |
| 13 | 0.224951 | 0.97437 | 0.230868 | 4.33148 | 77 |
| 14 | 0.241922 | 0.97030 | 0.249328 | 4.010784 | 76 |
| 15 | 0.258819 | 0.965926 | 0.267949 | 3.732054 | 75 |
| 16 | 0.275637 | 0.961262 | 0.286745 | 3.487418 | 74 |
| 17 | 0.292371 | 0.95630 | 0.30573 | 3.270856 | 73 |
| 18 | 0.309017 | 0.951057 | 0.324919 | 3.077686 | 72 |
| 19 | 0.325568 | 0.945519 | 0.344327 | 2.904214 | 71 |
| 20 | 0.34202 | 0.939693 | 0.36397 | 2.74748 | 70 |
| 21 | 0.358368 | 0.933581 | 0.383864 | 2.605091 | 69 |
| 22 | 0.374606 | 0.927184 | 0.404026 | 2.475089 | 68 |
| 23 | 0.390731 | 0.920505 | 0.424474 | 2.355855 | 67 |
| 24 | 0.406736 | 0.913546 | 0.445228 | 2.246039 | 66 |
| 25 | 0.422618 | 0.906308 | 0.466307 | 2.144509 | 65 |
| 26 | 0.438371 | 0.898794 | 0.487732 | 2.050306 | 64 |
| 27 | 0.45399 | 0.891007 | 0.509525 | 1.962612 | 63 |
| 28 | 0.469471 | 0.882948 | 0.531709 | 1.880728 | 62 |
| 29 | 0.484809 | 0.87462 | 0.554308 | 1.80405 | 61 |
| 30 | 0.5000 | 0.866026 | 0.57735 | 1.732053 | 60 |
| 31 | 0.515038 | 0.857168 | 0.60086 | 1.664281 | 59 |
| 32 | 0.529919 | 0.848048 | 0.624869 | 1.600336 | 58 |
| 33 | 0.544639 | 0.838671 | 0.649407 | 1.539867 | 57 |
| 34 | 0.559192 | 0.829038 | 0.674508 | 1.482563 | 56 |
| 35 | 0.573576 | 0.819152 | 0.700207 | 1.42815 | 55 |
| 36 | 0.587785 | 0.809017 | 0.726542 | 1.376383 | 54 |
| 37 | 0.601815 | 0.798636 | 0.753553 | 1.327046 | 53 |
| 38 | 0.615661 | 0.788011 | 0.781285 | 1.279943 | 52 |
| 39 | 0.62932 | 0.777146 | 0.809783 | 1.23490 | 51 |
| 40 | 0.642787 | 0.766045 | 0.83910 | 1.191755 | 50 |
| 41 | 0.656059 | 0.75471 | 0.869286 | 1.15037 | 49 |
| 42 | 0.66913 | 0.743145 | 0.90040 | 1.110614 | 48 |
| 43 | 0.68200 | 0.731354 | 0.932514 | 1.07237 | 47 |
| 44 | 0.694658 | 0.71934 | 0.965688 | 1.035532 | 46 |
| 45 | 0.707106 | 0.707107 | 1.00000 | 1.00000 | 45 |
| | cos | sin | | tan | |