15. STATISTICS AND PROBABILITY

STATISTICS

Statistics is a branch of mathematics that is used extensively in the real world to make decisions. Such decisions are made by gathering information, analysis and making predictions. Statistics equips us with tools to carry out an investigation in a systematic way and to arrive at a solution based on informed judgements.

In our daily lives, we are bombarded with problems to be investigated on a range of topics such as health, sports, student performance, opinions on issues, population trends, income growth and many other areas. The setting up of scientific experiments to tests hypotheses also rely on knowledge of statistics.

It is therefore essential to learn how to solve problems using the tools in statistics and probability. In this chapter, there will be an emphasis on various methods that are used to collect, organise, represent, interpret and analyse data, in order to make decisions. Statistics provides us with tools such as methods for organising and summarising data, representing data using graphs, tables and statistical diagrams. It is, therefore, a powerful form of communicating information in precise and visually appealing forms.

Sample statistics and population parameters

Supposing we wanted to investigate the attitude of Form 5 students to mathematics. It will indeed be desirable to ask every student in Form 5 about how they feel about mathematics. However, this approach would be time-consuming, costly, and logistically challenging. On the other hand, one can obtain information from a well-chosen subset of the Form 5 population and draw valid conclusions about how all students feel about mathematics.

The set of all students in Form 5 is an example of a **population**, while the subset of participants chosen to be interviewed is an example of a **sample.** If data is collected from every person (or object) in a population, we refer to it as a census.

Population parameters are measures associated with a population. This would include the mean score in mathematics (μ) and the proportion of students who like mathematics (p). Since it is almost impossible to collect data from all members of a population, these measures are not exact.

To obtain estimates of population parameters, we obtain **statistics** from a sample of the population. This

would include the mean score ($\bar{x}$) of a subset of the population in mathematics ($\bar{x}$) and the proportion of students who like mathematics ($\hat{p}$).

When we summarise, analyse and calculate measures of data from a sample, this is called **descriptive statistics.** When we use our sample data to make predictions about a population, such techniques are called **inferential statistics**. In this chapter, we will focus only on descriptive statistics. Inferential statistics involves knowledge of theories and principles in probability and is beyond the scope of this course.

Types of variables

The type of data we collect will depend on the characteristic we are investigating. We may be interested in the mass of a new-born baby, test scores, type of transport or the opinion of persons on a social issue. Each characteristic is associated with a variable to be investigated. A **variable** is any characteristic whose value may change from one object to the other in a population. Variables can be **qualitative** (non-numerical) or **quantitative** (numerical).

Qualitative variables are usually categorical in nature. Some examples of categorical data are shown below. Gender, for example, has only two possible categories and is called a binary variable. Other variables have at least three categories.

Qualitative Variable	Categories
Gender	Male, Female
Ice Cream Flavours	Vanilla, Cherry, Soursop
Type of transport	Car, Bus, train
Budget expenditure	Food, Rent, Transport
Month of the year	January, February, March
Make of car	Toyota, Nissan, BMW
Colour of T-shirts	White, Blue, Green, Red

In quantitative data, the value of the characteristic is a number, for example, the number of children in a family, shoe size or the height of a person.

Numerical data can be of two types. It is **discrete** if it can only take particular values within an interval. Examples of discrete data are given in the table below. Notice that there are no values between any two consecutive values.

Discrete Variable	Possible Values
Number of items correct on a test	0,1,2,3, …
Shoe size	2, $2\frac{1}{2}$, 3, $3\frac{1}{2}$, 4, $4\frac{1}{2}$,…
The number of groceries in a city	30, 32, 60…
The number of skittles in a pack	50, 51, 53, …
Floor number in a building	1, 2, 3, …
Ranking of product on its quality	1, 2, 3, …

A continuous variable has an infinite number of possible values between any two points on the measurement scale. Measures of distance, weight, time and temperature are examples of continuous variables. Some of these are shown below.

Continuous Variables	Possible Values
The height of seedlings in a sample (cm)	4.5, 5, 6.1,7.3. 8.0
The temperature of a mixture (^{0}C)	45.23, 46.45, 47.78, 50.04
The time taken to run 100m distance (secs)	10.369,10.258, 10.698,
The mass of sugar (g) in a serving of cake	3.2, 4.5, 5.1, …
The speed of a vehicle in km/h	60, 70, 80, …

With continuous variables, the exact value of the measure is not possible to determine. Our degree of accuracy depends on our measuring device. For example, mouse weight will have an infinite number of possible values between 25 grams and 26 grams because one could always add extra decimal places to the measurement.

While it is possible to summarise, represent and analyse data collected on variables that are either quantitative and qualitative, it must be noted that the techniques used for qualitative data will differ from those for quantitative data (discrete and continuous).

To explore these differences more closely, we consider the following example comprising data collected on five variables from a group of 40 students in a graduating class.

Gender	Meal Preference	Hat size	Gown Size	Height (cm)
F	Vegetarian	XS	28	150
M	Chicken	S	36	161
M	Chicken	L	37	170
F	Fish	Med	35	166
F	Vegetarian	S	29	160
F	Chicken	Med	36	163
F	Fish	Med	35	162
M	Fish	XS	29	165
M	Fish	XL	42	168
F	Chicken	S	30	161
M	Chicken	S	31	175
M	Chicken	S	30	170
M	Fish	L	38	173
F	Chicken	XS	28	152
F	Fish	XS	33	151
F	Vegetarian	S	30	156
M	Chicken	Med	32	169
M	Chicken	L	38	177
M	Chicken	XS	29	160
M	Chicken	S	36	158
F	Chicken	Med	32	165
F	Fish	Med	33	167
F	Chicken	Med	31	159
M	Fish	XL	41	172
F	Fish	XS	29	151
M	Vegetarian	Med	36	175
M	Chicken	L	37	179
F	Fish	S	27	155
F	Fish	S	30	155
F	Chicken	XS	28	150
M	Fish	Med	35	169
M	Fish	XL	40	175
M	Chicken	Med	35	172
M	Fish	Med	33	170
F	Fish	L	40	170
F	Chicken	XL	42	168
M	Vegetarian	S	29	174
F	Fish	S	32	155
F	Chicken	XS	30	159
F	Chicken	Med	34	166

Table 1: Data on graduating class

M – Male	S – Small	Med – Medium
F – Female	XS – Extra Small	L – Large
		XL – Extra Large

We summarise variables by type as follows:

Variable	Type	Values
Gender	Qualitative Categorical	{M, F}
Gown size	Qualitative Categorical	{XS, S, Med, L, XL}
Meal Preference	Qualitative Categorical	{Veg, Chicken, Fish}
Height	Quantitative Continuous	$150 \leq x \leq 179$
Ring size	Quantitative Discrete	{3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12}

Frequency distributions

Data that is not processed or organised is called raw data. The data shown in the above table is raw data collected on four variables. It is difficult to draw conclusions from raw data by simply looking at the values. A common method of organising raw data is to construct a frequency distribution. This is usually the first and simplest step in analysing data.

Ungrouped Frequency Distributions

A frequency table is used to summarise a set of data. It records the *frequency* or number of times each value occurs. There are usually two columns in the table. The first column lists the set of possible scores, while the second column records the frequency of each score.

Let us first consider the data on gender. Since there are only two categories we can count the number of males and deduce the number of females by subtracting the total from 40. We can summarise it a frequency table.

Gender	Frequency
Male (M)	19
Female (F)	21
TOTAL	40

Table 2

We can construct a similar frequency distribution for meal preference.

Meal Preference	Frequency
Veg	5
Chicken	19
Fish	16
TOTAL	40

Table 3

To summarise the data on hat size, we list the categories in one column and use tally marks to record the entries. This will ensure that we do not count an entry twice. Starting from the top, we record a tally mark in the frequency column next XS, then one next to S, then L and so on until all forty marks are recorded.

Notice that the tally marks are in bunched in groups of five because it is easy to count in fives.

The data on hat size is summarised below.

Hat Size	Tally	Frequency
XS	⊞ ⊞ |||	8
S	⊞ ⊞ |	11
Med	⊞ ⊞ ||	12
L	⊞	5
XL	||||	4
TOTAL		40

Table 4

Grouped Frequency Distributions (discrete data)

The data on gown sizes has 15 possible values and if represented in an ungrouped distribution, there will be 15 rows. This will not permit meaningful results and is better to group the data so as to reduce the number of rows. We can select five classes each containing 3 values. These are:

$$28 - 30 \quad = \{28, 29, 30\}$$
$$31 - 33 \quad = \{31, 32, 33\}$$
$$34 - 36 \quad = \{34, 35, 36\}$$
$$37 - 39 \quad = \{37, 38, 39\}$$
$$40 - 42 \quad = \{40, 41, 42\}$$

Gown size	Frequency
28 – 30	8
31 – 33	12
34 – 36	11
37 – 39	5
40 – 42	4
TOTAL	40

Table 5

The data on height has a range from 150-177 and should be grouped as well. We can choose to make groups with 5 possible values each. So, the first group will comprise all the scores in the class interval 150-154, the other class intervals will be 155-159, 160-164, 165-169, 170-174 and 175-179.

Grouped Frequency Distributions (discrete data)

In tabulating grouped data, we form classes or groups to accommodate all the raw data. The **class limits** specify the range of data values that fall within a class**.**
Class limits are defined as the minimum and the maximum values the class interval contains. The minimum value is known as the lower-class limit (LCL) and the maximum value is known as the upper- class limit (UCL). For the class 150-154, the LCL is 150 and the UCL is 154.

It was mentioned earlier that for continuous data, the true value of the measure is not really known due to lack of precision in measuring instruments. For example, the class 150 – 154 will have heights rounded to the nearest cm and can actually contain values from 149.5 and up to but not including 154.5. The class 155-159 will have heights from 154.5 and up to but not including 159.5. These end points are the **class boundaries.** In the first interval, 149.5 is the lower-class boundary and 154.5 is the upper-class boundary. These class limits and class boundaries for the data on the height of students are shown in the table.

Height of student (cm)		Frequency
Class limits	Class boundaries	
150-154	$149.5 \leq x < 154.5$	6
155-159	$154.5 \leq x < 159.5$	5
160-164	$159.5 \leq x < 164.5$	7
165-169	$164.5 \leq x < 169.5$	9
170-174	$169.5 \leq x < 174.5$	8
175-179	$174.5 \leq x < 179.5$	5
TOTAL		40

Table 6

Class boundaries take into account errors in measurement and are appropriate for summarising continuous data. On the other hand, class limits do not cater for errors in measurement and are quite appropriate in summarising discrete data.

Displaying data

When data is collected and organised into frequency tables, it is easy to summarise, interpret and make decisions. However, displaying the data using charts is an even more effective means of organising it. There are various types of charts and these can be drawn by hand or using computer programmes, such as EXCEL.

In displaying data, we must consider whether the data is discrete or continuous. This will influence our choice of the chart. In addition to the type of data, the purpose of the display is also an important consideration when making choices about the type of chart to select. in the following section, we will display the data on characteristics the characteristics of the 40 students using a few common charts.

Pie charts

Pie Charts or Circle graphs are useful when we wish to make simple comparisons or show how parts compare with a whole. It is usually used for representing categorical data when the number of categories is relatively small.

When data is represented on pie charts, the sizes of the sectors are proportional to the categories sampled. Using the data on gender in Table 2, the following example shows how a pie chart is constructed.

Category	Frequency	Angle in sector
Male	19	$\frac{19}{40} \times 360^0 = 171^0$
Female	21	$\frac{21}{40} \times 360^0 = 189^0$
Total	40	
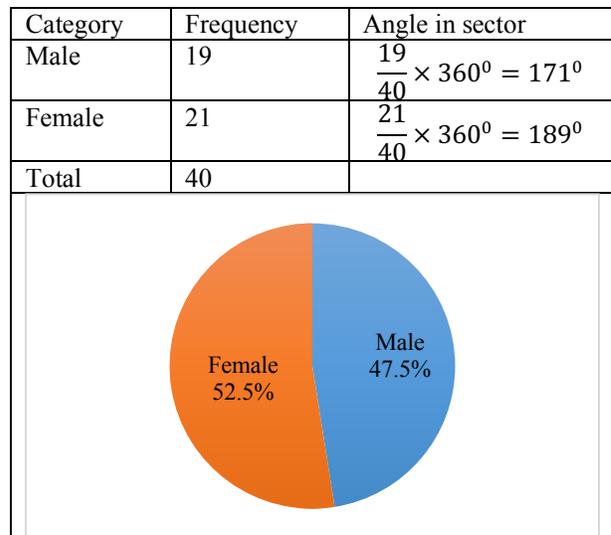
Bar charts

Bar graphs are used to display data when the variables are discrete. Variables can be categorical or numerical. The Bar graph has the following characteristic features:

1. The spaces between the bars are equal.
2. The widths of the bars are the same.
3. The height of the bar gives the frequency.
4. The bar chart is best drawn on a grid.
5. The bars can be drawn horizontally or vertically

The data on meal preference (Table 3) and hat size (Table 4) is displayed below.

## Meal Preference of Students

Hat size of students
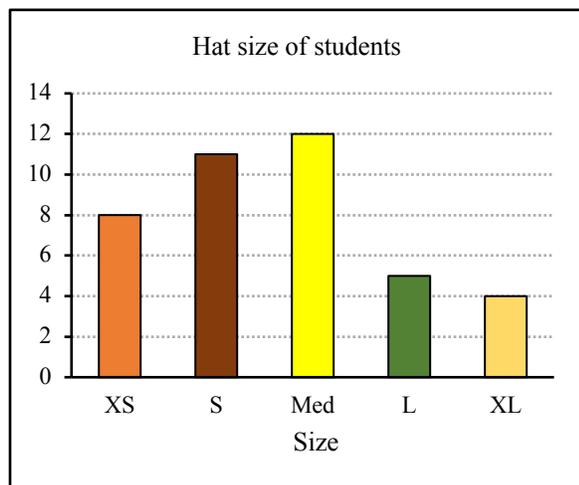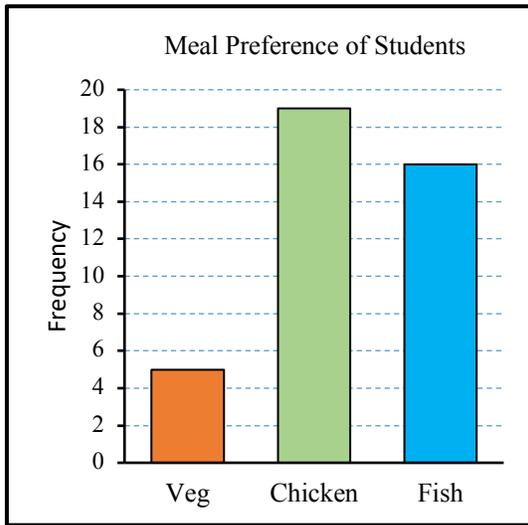
Histograms

Histograms are used to display continuous data and are useful in showing the shape of a distribution. It resembles a bar graph except that there are no 'gaps' between the bars. This is because the histogram represents data in which the variables are continuous and as such, the variables can take any value in the interval on the horizontal axis. So, where one bar ends the other starts.

In constructing a histogram, upper-class boundaries are plotted against the frequencies. In the histogram shown below, all the bars are equal in width and their height is equal to the frequency. Histograms can have bars that are unequal in width and in such a case the

area of the bars is equal to the frequency. This is best when one bar is too tall with respect to the other. In such a case, the width may be doubled and the height halved so as to maintain the same area.

The Histogram below shows the data in Table 5, the heights of the 40 students in the sample.

Height of students
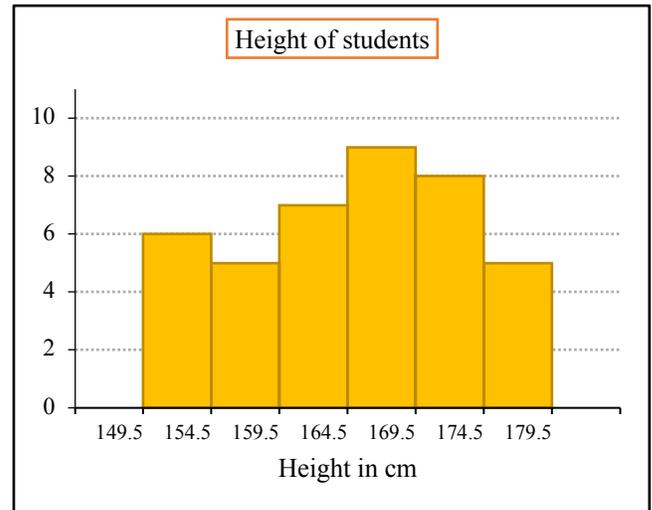
Frequency Polygons

If the mid-points of the bars of a histogram are joined by straight lines a frequency polygon is formed. Like the histogram, the frequency polygon is useful in displaying the shape of a distribution. In a frequency polygon, we plot the midpoints of the class intervals against the frequencies.

Using the data on gown size in Table 3, we will now draw a frequency polygon. First, we must obtain the mid-point of each class interval. We calculate these using ½ (Upper limit + Lower limit).

To close the polygon, we extend the table to include one unit with zero frequencies at the two ends. Notice, we checked backwards to obtain (26, 0) and (44, 0) so that the polygon is closed by starting and ending on the horizontal axis.

Gown size	Mid-Point	Frequency
	26	0
28 – 30	29	8
31 – 33	32	12
34 – 36	35	11
37 – 39	38	5
40 – 42	41	4
	44	0

The completed frequency polygon is shown below.

125

Distribution of gown sizes
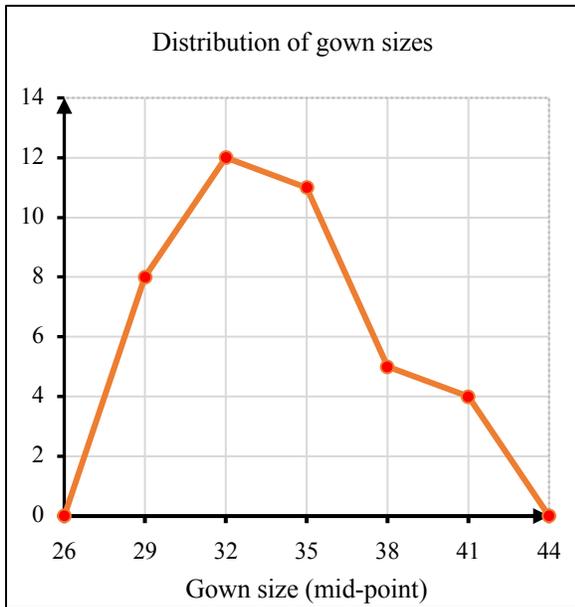
Line graphs

Line graphs are useful means of displaying statistical data when we wish to examine trends or growth over a period of time. It is used extensively in business, where companies may wish to predict growth in income or sales over a period of time. In medicine, physicians may use these graphs to monitor fluctuations in blood sugar levels, blood pressure and temperature.

Consider the following data:

A school recorded the profits over the period 2012-2017 based on sales in the cafeteria. The results are shown in the table below.

Year	Profit
2012	10 000
2013	12 000
2014	13 000
2015	10 000
2016	15 000
2017	18 000

We wish to represent this data on a line graph and comment on the trend if any.

In constructing a line graph, the following should be noted:
1. The time (year, month, day etc.) is always on the horizontal axis.
2. Consecutive points are joined by straight lines.
3. To predict future values, we draw a trend line.

Profit over the period 2011-2017
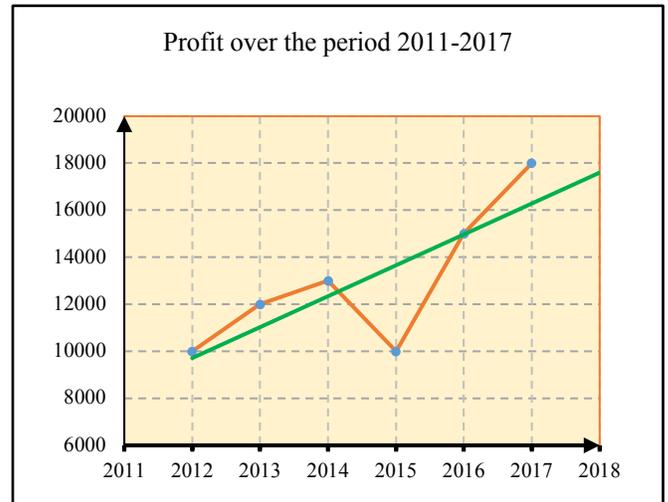
Interpreting a Line Graph

The line graph shows that there was an increasing trend over the period 2011-2018, except in 2015 where there was a decrease in profits.

A useful feature of the line graph is that it allows us to predict future values, with a reasonable degree of accuracy, if there is a trend. To do so, we draw a *trend line*, which is a line of *best fit* or the best possible straight line for the data. The trend line is shown in the above diagram. This line is drawn by finding a straight line that minimizes the vertical displacements from the plotted points. It can be done manually, but EXCEL can generate it automatically.

As shown in the figure above, the trend line has a positive slope indicating an increasing trend. By producing it until it is in line with the next point on the horizontal axis, the profits for 2018 is predicted as just under $18 000. Of course, this value is only an estimate and it is based purely on past profits. It also assumes that conditions will remain the same in the next year. If conditions change, then the estimate will not be accurate.

This procedure of predicting future values using a trend line is called **extrapolation**. The line graph can be used to predict future values or in-between values, provided that the data follows a definite trend. When such predictions are made, an assumption is made that conditions remained the same. Line graphs are also used for **interpolation** which is obtaining a value between two consecutive values.

In the line graph shown below**, t**he temperature of a hot drink was recorded every two minutes after

placing a block of ice in it. To predict the temperature of the drink after 7 minutes.

Cooling of a hot drink using ice
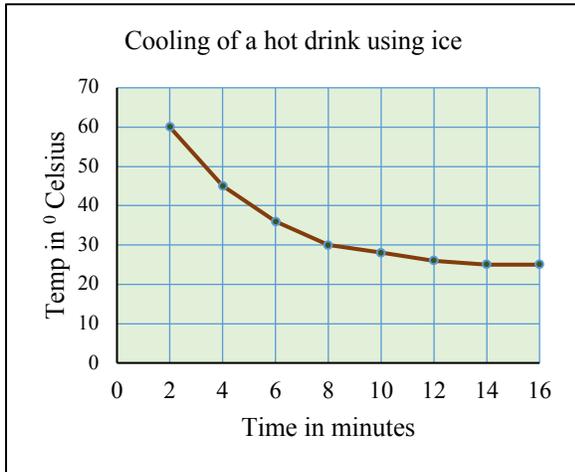
After 7 minutes, the predicted temperature is approximately 32^0 (by read-off). This procedure of predicting in-between values using a line graph is called **interpolation**.

Example 1

The marks obtained by 50 students in a quiz out of five questions are displayed in the table below. Draw a Bar Graph to represent the data.

Marks	0	1	2	3	4	5
Frequency	1	4	7	10	18	10

Solution

Marks of students in a Quiz
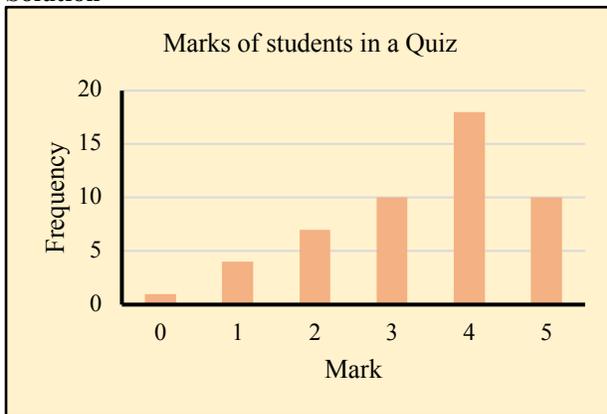
In this example, the data was numerical, but discrete so a Bar Graph is appropriate.

Example 2

The following data represents the distance in kilometers between 60 towns from their local airport.

21	10	30	22	33	5
37	12	25	42	15	39
26	32	18	27	28	19
29	35	31	24	36	18
20	38	22	44	16	24
10	27	39	28	49	29
32	23	31	21	34	22
23	36	24	36	33	47
48	50	39	20	7	16
36	45	47	30	22	17

(a) Draw a frequency table to represent this data.
(b) Draw a Histogram to display the data.
(c) Draw a frequency polygon to display the scores and comment on the shape of the distribution.

Solution

(a) The data has a large spread ranging from 5 to 49. In this case, we group the scores so that the table is easy to interpret.

Score	Frequency
1 - 10	4
11 - 20	10
21 - 30	21
31 - 40	17
41 - 50	8
Total	60

(b) To draw the Histogram, we calculate the class limits since the scores are continuous.

Score		Frequency
Class limits	Class boundaries	
1 - 10	$0.5 \leq x < 10.5$	4
11 - 20	$10.5 \leq x < 20.5$	10
21 - 30	$20.5 \leq x < 30.5$	21
31 - 40	$30.5 \leq x < 40.5$	17
41 - 50	$40.5 \leq x < 50.5$	8
Total		60

(b) The Histogram is drawn below.
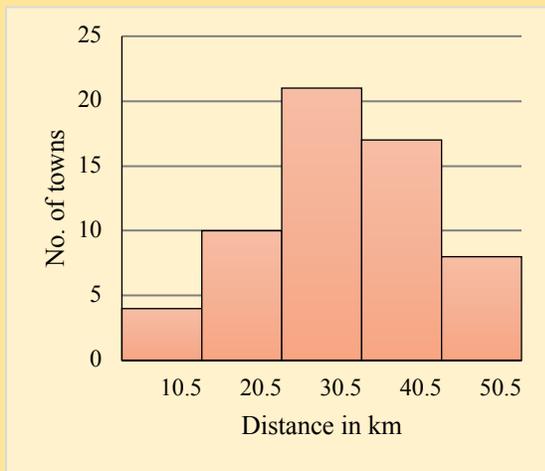
(c) The mid-points are first calculated.

Distance	Mid-Point	Frequency
	0	0
1 − 10	5.5	4
11 − 20	15.5	10
21 − 30	25.5	21
31 − 40	35.5	17
41 − 50	45.5	8
	55.5	0

The Frequency Polygon is drawn below.
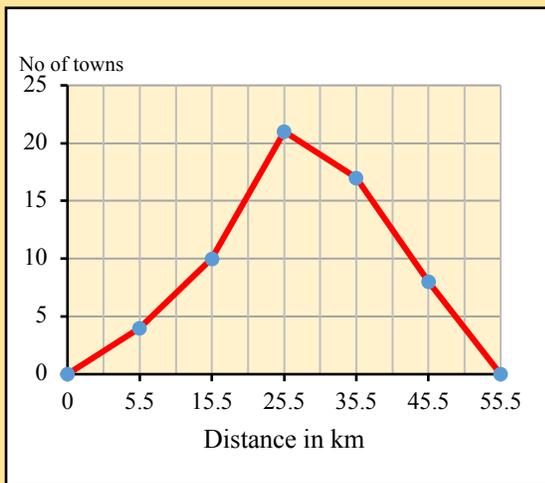
The shape of the distribution indicates that the majority of towns are between 15-35 km from the airport. A few are further away and fewer are close to the airport.

STATISTICAL INDICES

So far, we have been looking at how data can be collected, summarised in tables and how they can be displayed and interpreted in graphs. There are many other types of graphs for displaying data but these are not covered at this level.

Data can also be summarised using statistical indices such as the mean, median and mode. These indices require mathematical computations and researchers usually assign numbers to variables so as to facilitate computation. In performing such computations on statistical variables, we need to consider the nature and limitations of the measurement scales used to quantify the data. Knowing the level of measurement helps one to decide what statistical analysis is appropriate on the values that were assigned.

Measurement Scales

The type of unit on which a variable is measured is called a scale. Statisticians refer to four types of measurement scales or four levels of measurement. These are (1) nominal, (2) ordinal, (3) interval, and (4) ratio.

Nominal Scale

When we assign numbers to qualitative variables such as gender, religion or meal preference we are using a nominal scale. Variables on a nominal scale are often categorical variables.

The key feature of a nominal scale is that there is no inherent quantitative difference among the categories. Nominal scales are used to identify and differentiate between categories. For example, a researcher may assign the numbers such as:

Male	1
Female	2

Vegetarian	– 5
Fish	– 6
Chicken	– 7

The numbers are arbitrarily chosen and there is no ordering of categories (no category is better or worse, or more or less than another). The nominal scale is lowest in level and only serves the purpose of identification.

When analysing data that is nominal in nature, we can only tally the frequencies and compare categories by counting. Bar graphs and Pie Charts are appropriate for displaying the data. We can also determine the mode and median of nominal variables.

Ordinal Scale

Sometimes we assign numbers to rank observations based on a performance, or dress size. Ordinal scales are used for categorical variables. Examples of ordinal scales are shown below.

First – 1	Extra Small– 1
Second – 2	Small – 2
Third – 3	Medium– 3
	Large– 4
	Extra Large– 5

In addition to the identification of the categories, ordinal scales have the property of magnitude. In other words, the order is important. However, the distance between scale points is not equal. For example, the difference between 1 and 2 is not the same as the difference between 2 and 3. Also, we cannot say that the 1st rank is 3 times more than the 3rd rank in the attribute measured. Think of a beauty contest - the winner is rated as better than those who came second or third, but we cannot quantify how much better the winner was than the runners-up.

Research in the social sciences use questionnaires to collect data on opinions of respondents and do so using a 5-point Likert scale ranging from Strongly Disagree to Strongly Agree (SD – 1, D – 2, Neutral – 3, A – 4, SA – 5). Most researchers consider Likert scales as ordinal scales because one cannot say that the 1- point difference between "Strongly Agree" and "Agree" represent the same amount of change in the strength of an attitude as the 1-point difference between "Agree" and "Neutral".

Ordinal data, therefore, tells us more than nominal data but it does not tell us about differences on a scale and we should not add or subtract scores. We can compare frequencies by tallying scores or using bar graphs and pie charts to display categories.

Interval Scales

An interval scale has all the characteristics of a nominal and an ordinal scale but, in addition, it is based on equal intervals. In such a scale, the difference between 1 and 2 is the same as the distance between 2 and 3. Interval scales are used with continuous variables and permit more complex statistical calculations. Examples of interval scales are rating scales, where respondents are asked to rate the quality of an item on a scale from 0 to 10. Test scores are also examples of interval scales. Candidates can score from 0 to 100% and there are equal intervals between scores.

The advantage of interval scales over ordinal scales is that the difference between a rating of 6 and 7 is the same as the difference between a rating of say, 2 and 3. However, we cannot say that a rating of 5 is five times as strong as a rating of 1. Also, a rating of zero is arbitrary and does not specify the absence of the attribute. Another example of an interval scale is temperature, measured either in degrees Celsius or degrees Fahrenheit. The intervals are equidistant (i.e., a 1-degree increase from 15 to 16 degrees represents the same amount of increase in temperature as does 30 to 31 degrees, but 30 degrees is not "twice as hot" as 15 degrees).

Although interval scales permit one to add, subtract and find averages of scores, there are some limitations when compared to ratio scales.

Ratio Scales

Ratio scales are the highest and most precise level of measurement. Measures of mass, time, distance and speed are examples of ratio scales. Unlike interval scales, ratio scales have an absolute zero. For example, if a candidate scores zero on a spelling test, it does not mean that the candidate cannot spell a single word. In other words, the score of zero is relative to what was tested. It is not an absolute zero.

Ratio scales are the only ones that have a multiplicative property. It is possible to have zero mass, length or speed (although in the real world this may never be observed). Because of this property, it is possible to make statements that 30 mg of medicine is three times as strong as 10 mg. One can perform any of the four arithmetical operations on measures that have ratio scales.

If there is a choice among measurement scales, then we always select the scale with the highest level. Hence, an ordinal scale should be preferred to a nominal scale, an interval scale to an ordinal scale, and a ratio scale to an interval scale.

MEASURES OF CENTRAL TENDENCY

In statistics, we are often required to summarise a set of scores by obtaining a single score to represent the entire distribution. Measures of central tendency are one such measure and there are three of these – the mean, the median and the mode. Each measure is a typical central value that represents all the scores in the distribution. However, each measure has its unique properties and contributes separately to the overall understanding of the data set.

The mean

The mean of a set of scores is the sum of a set of scores divided by the number of scores in the set. The mean is another name for what is commonly called the *arithmetical average*.

When data is presented in the raw form, we can calculate the exact value of the mean by simply adding up the scores and dividing the total by the number of scores.

Example 3

Calculate the mean of the set of scores:
6, 7, 8, 10, 12, 13, 15, 16, 20 and 20

Solution

The mean is calculated as

$$\bar{X} = \frac{6+7+8+10+12+13+15+16+20+20}{10} = 12.7$$

where the symbol, $\bar{X}$, pronounced X Bar, represents the mean.
If the variable, x, represents a score, then the formula for mean, is written as $\bar{X} = \frac{\sum x}{n}$.

In the formula, the symbol $\sum$ means 'the sum of' and n represents the number of scores.

Example 4

On five Chemistry tests, John received the following scores: 72, 86, 92, 63, and 77.
   i.   What is his mean score on the five tests?
   ii.   What test score must John earn on his sixth test so that his mean score for all six tests will be 80?

Solution

   i.   Mean score on 5 tests

$$= \frac{\sum x}{n} = \frac{72+86+92+63+77}{5} = \frac{390}{5} = 78$$

   ii.   Since his average on all 6 tests is to be 80, then the sum of all scores on the six tests is $\sum x = 80 \times 6 = 480$.

The total must include the sum of scores on the first 5 tests, together with the score on the sixth test,
The score on the 6th test should be
$= 480 - (72+86+92+63+77) = 480 - 390 = 90$

Properties of the mean

An important property of the mean is that it includes every value in the data set, as a part of the calculation. In addition, the mean is the only measure of central tendency where the sum of the deviations of each value from the mean is always zero. This is explained in the following example.

Five (5) students collected money on a donation sheet, to raise funds for a school project. The amount collected by each student is $50, $60, $65, $75, $80.

The mean amount of money collected in dollars is calculated to be

$$\bar{X} = \frac{\sum x}{n} = \frac{50+60+65+75+80}{5} = \frac{330}{5} = 66$$

Since we took the total collected by all 5 students and divided it by 5, this is equivalent to dividing the total donations equally among the five students. We can think of $66, as representing the amount of funds collected by each student, if each had collected the same amount. In this sense, the mean 'evens off' all the scores to a single value. Each score (x) differs from the mean score by a value called the deviation, $(x - \bar{X})$. These deviations are computed and shown in the table.

Student	A	B	C	D	E
$x(\$)$	50	60	65	75	80
$(x - \bar{X})$	50−66	60−66	65−66	75−66	80−66
$(x - \bar{X})$	−16	− 6	−1	9	14

If we were to sum the deviations we obtain a total deviation of zero. This is because students A, B and C had scores below the mean (a total deviation of −23), while students, C and D had scores above the mean (a total deviation of +23).

This property of the mean is extremely important and the mean is used in computing many other statistical indices. The mean, however, has one disadvantage. This is illustrated in the following example.

A group of 10 students recorded the number of books they read during the Easter vacation. Their scores are 3, 2, 5, 1, 4, 4, 3, 2, 0, 23.
The mean score,

$$\bar{X} = \frac{3+2+5+1+4+4+3+2+0+23}{10} = 4.7.$$

If this was a truly typical score, it would seem that on the average 4-5 books were read by each person. However, the data does not show this to be true because except for one student, everyone read 4 or less than 4 books. The extremely large score of 23, made by one student, influenced the mean score.

In this case, the mean is not a true representative of the group and it would be misleading to use it as a measure of central tendency. There are other measures that would give a more reasonable picture. The median is one such measure.

The median

Another measure of central tendency that is used to represent a set of scores is the median. The median of a set of scores is the middle value when the scores are arranged in either ascending or descending order of magnitude. If there are two middle values, then the median is the mean of the middle two values.

Example 5

Determine the median of the following set of scores:
4, 5, 7, 3, 9, 0, 6, 8, 7

Solution

First, arrange the scores in either ascending or descending order.
0 3 4 5 6 7 7 8 9
Median

The score of 6 is the middle score as it divides the set into two equivalent sets with four members on each side.

Example 6

Determine the median of the following set of scores:
8, 19, 15, 16, 23, 19, 10, 16, 18, 17, 19, 12

Solution

First, arrange the scores in either ascending or descending order.
Two scores, 16 and 17, are in the middle. The median is the mean of both scores.

8 10 12 15 16 16 17 18 19 19 19 23

Median = (16 + 17) ÷ 2 = 16.5

If we have a large number of scores, it will not be possible to list them all and find the middle. We can use a simple formula to calculate the rank of the median.

For a set of n scores, the middle score is the $\left(\dfrac{n+1}{2}\right)th$ score.

When n is odd, this works out to be a single score.

For example, when $n = 41$, the median is the $\dfrac{41+1}{2} = 21$st score.

When n is even, this works out to be the mean of two scores.

For $n = 40$, the formula gives $\dfrac{40+1}{2} = 20.5$.

This means that the median is halfway between the 20th and the 21st score. The median is obtained by calculating the mean of the 20th and the 21st scores.

Properties of the median

The median always gives the center of a distribution, even when there are extreme scores. To find the median of the scores in the example where the number of books read by a set of 10 students was:

3, 2, 5, 1, 4, 4, 3, 2, 0, 23.

We first arrange the scores in ascending order to determine where the median will lie.

0 1 2 2 3   3 4 4 5 23

Median is (3 + 3) ÷ 2 = 3

In this example, there is no middle score (the number of scores is even), so the median lies between the 5th and 6th scores. The median is therefore 3.

We can interpret this as exactly one-half of the students read less than 3 books and half of them read more than 3 books. The extreme score, 23 did not affect the median. In this case, the median gives a more realistic representation of the center of the distribution than the mean score of 4.7.

The Mode

The mode is the value (number) or category that occurs the most or is the more frequent occurrence in a distribution.

Example 7

Determine the mode of the set of scores:
8, 10, 12, 15, 16, 16, 17, 18, 19, 19, 19, 19, 19, 19

Solution

Clearly, the score 19 occurs the most frequently. Hence, 19 is the mode.

Example 8

Determine the mode of the set of scores:
17, 7, 3, 12, 3, 3, 5, 5, 12, 10, 5, 24

Solution

Since the scores 3 and 5 both occur 'the most frequently', there are two modes. These are 3 and 5.

Properties of the mode

One can usually obtain the mode of a distribution by mere inspection. No calculations are required to do this and so the mode has the advantage of being easy to obtain. It is a widely used statistic in marketing because suppliers are interested in what products are in demand. A shoe store needs to know the more popular shoe sizes so that they can order these in the largest quantities. A supermarket owner would need information on brands of items that are sold in the community to avoid wastage and also to replenish supplies.

Unlike the mean and median, the mode is not necessarily a numerical score and is used in the analysis of categorical data. For example, we may refer to the modal form of transport. It is also possible to have more than one mode and even no mode in a distribution.

Mean, median and mode from frequency tables

Sometimes we are not given the raw scores, but a set of scores arranged in a frequency table. All three measures can be determined when data is organised in tabular form.

Mean from ungrouped frequency distributions

The scores obtained from tossing a die a total of 40 times are recorded in the table below.

Score	1	2	3	4	5	6
Frequency	8	5	9	6	7	5

To obtain the mean score, we need to calculate the sum of all 40 scores. Since the frequency tells us how many times each score occurs, we can obtain the total

for each of the individual scores by multiplying each score by its frequency. Rearranging the table in vertical form, we insert a third column to record these totals under the heading, $f \times x$.

Score (x)	Frequency (f)	$f \times x$
1	8	8
2	5	10
3	9	27
4	6	24
5	7	35
6	5	30
	$\sum f = 40$	$\sum fx = 134$

Next, we need to calculate two important totals.

1. The sum of the values in the **frequency** column –this gives the **number of scores**
2. The sum of the values in the $f \times x$ column- this gives the **sum of all the scores**.

In this example, there are **40** scores whose total is 134.
To compute the mean, we now substitute our totals in the formula

$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{134}{40} = 3.35$$

Mean from grouped frequency distributions

When data is grouped, it is not possible to reconstruct the raw data. Hence, we cannot obtain an exact value for the mean. However, it is possible to obtain a very good estimate of the mean using the following method. Consider the following data in which the masses of peas were recorded.

Mass in grams	No. of peas (f)
3 – 7	3
8 – 12	8
13 – 17	12
18 – 22	10
23 – 27	7
	40

We know that 3 peas had masses that were between 3-7 grams. This means that their masses could have been 3, 4, 5, 6, or 7 grams. To obtain the total mass of the 3 peas in this interval, we assume that their

average mass is 5 grams and multiply 3 by 5 to obtain a total mass of 15 grams.

We make an assumption that on the average, the set of values in an interval will approximate to the midpoint of the interval. In this case, 5 is the mid-point of the interval 3-7. We now treat the midpoint as the variable score, x, and proceed to use the same formula as was done for ungrouped data. The computations are shown in the next table.

Mass in grams	Mid-Point (x)	No. of peas (f)	$f \times x$
3 – 7	5	3	15
8 – 12	10	8	80
13 – 17	15	12	180
18 – 22	20	10	200
23 – 27	25	7	175
		40	650

Mean mass, $\overline{X} = \dfrac{\sum fx}{\sum f} = \dfrac{650}{40} = 16.3$ grams

It should be noted that this method would give rise to an estimate of the mean since the actual raw scores were not available.

> The mean, $\overline{X}$ from a frequency table, is $\overline{X} = \dfrac{\sum fx}{\sum f}$ where x represents the mid-point of the class interval and f the frequency of x.

Median from ungrouped frequency distributions

Using the data on the scores obtained from tossing a die a total of 40 times, let us now obtain the median.

Score	Frequency
1	8
2	5
3	9
4	6
5	7
6	5
	Total = 40

We wish to determine the middle score in a set of 40 scores. Since the number of scores is even, there are two scores in the middle scores. To obtain the rank of the middle score, we can line up the 40 scores in ascending order.
We would obtain the following list:

1 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 …

The 20th score is 3

The 20th score is 3 and the 21st score is also 3. Hence the median score is 3.

[Alternatively, the middle score is $\left(\dfrac{n+1}{2}\right)th$

So, $(40 + 1) \div 2 = 20.5$, the two middle scores are the 20th and 21st scores.]

When the data is grouped, the exact value for the median cannot be obtained. We can estimate this value using a cumulative frequency curve (see section on measures of spread).

Mode from frequency distributions

When data is presented in frequency tables, it is easy to determine the mode. The score that occurs the most will have the highest frequency. Consider the following frequency tables for grouped and ungrouped data.

Score	Frequency
1	8
2	5
3	9
4	6
5	7
6	5
	40

Mass in grams	No. of peas (f)
3 – 7	3
8 – 12	8
13 – 17	12
18 – 22	10
23 – 27	7
	40

For the ungrouped data, the score 3 has the highest frequency of 9, indicating that it occurred most often. The mode is therefore 3.

For the grouped data, the class interval 13-17 has the highest frequency of 12, indicating that scores in this interval occurred most often. The modal class is therefore 13-17. The mid-point, 15 is an estimate of the mode

MEASURES OF SPREAD OR DISPERSION

If a sample has all its scores close together in value, then there is little variability; we say that the scores are homogenous. If the scores are widely dispersed,

there will be considerable variability and we say that the scores are non-homogenous. This variation between values in a distribution is called spread or dispersion.

There are three measures of spread- the range, the semi-interquartile range and the standard deviation.

The Range

The range of a distribution is the crudest measure of the spread. It is the difference between the largest and the smallest observed value in a distribution.

Example 9

Calculate the range for the data set: 75, 73, 88, 56, 73, 52, 11, 56

Solution

The range is 88 - 11 = 77.

Example 10

For data in a frequency table, determine the range.

Mass in grams	No. of peas
3 – 7	3
8 – 12	8
13 – 17	12
18 – 22	10
23 – 27	7

Solution

The smallest and largest scores are taken at the mid-points of the end classes.
Hence, the range is: $25 - 5 = 20$ grams.

Quartiles

The range ignores the remaining data in a distribution and considers only the two end scores. It is greatly influenced by the presence of just one unusually large or small score in the sample. As such, we need other measures to measure dispersion.

Quartiles divide the set of scores into four equal sets. The three quartiles are called the lower quartile, denoted by Q_1, the middle quartile, which is the median, denoted by Q_2 and the upper quartile denoted by Q_3.
Consider the raw scores:
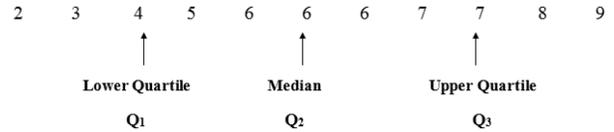
For a larger number of scores, it is best to use a formula to determine the rank of the quartiles.
For a set of n scores, let the location of the ranks of the quartiles be L_{Q_1}, L_{Q_2} and L_{Q_3} respectively.
Then,

$$L_{Q_1} = \frac{(n+1)}{4}, \ L_{Q_2} = \frac{(n+1)}{2} \text{ and } L_{Q_1} = \frac{3(n+1)}{4}$$

Semi Interquartile range

The inter-quartile range is a measure of the spread or dispersion. It is calculated by taking the difference between the **upper** and the **lower** quartiles. The Inter Quartile Range (IQR) is the width of an interval that contains the middle 50% of the sample. By ignoring the lower and upper scores, is not affected by extreme scores. Hence, it is a better measure of a spread than the range.
The Inter-Quartile Range
$= Q_3 - Q_1$
The Semi Inter Quartile Range is one half of the IQR.
$= \frac{1}{2}(Q_3 - Q_1)$

Example 11

Determine the inter-quartile range and the semi-interquartile range for the data:
4, 4, 5, 5, 6, 7, 8, 9, 10, 11, 13, 15, 16, 18, 19, 20

Solution

The data is already ordered and there are 16 scores. We first divide the scores into two equal sets and then find the median of each set.

4 4 5 5 ┊ 6 7 8 9 ┊ 10 11 13 15 ┊ 16 18 19 20

The median is mid-way between 9 and 10, so Q_2 = 9.5
The lower quartile is mid-way between 5 and 6, so Q_1 = 5.5
The upper quartile is mid-way between 15 and 16, so Q_3 = 15.5
The Interquartile Range is: $Q_3 - Q_1$ = 15.5 – 5.5 = 10
The Semi-Interquartile Range is: ½ ($Q_3 - Q_1$) = ½ (15.5 – 5.5) = 5

Median and quartiles from cumulative frequency curve

To determine the median and quartiles from a grouped frequency table, we use a cumulative frequency curve. Consider the data on scores of students in a test out of 50 shown in the table below. We can obtain estimates of these measures using a cumulative frequency curve (Ogive).

Before drawing the cumulative frequency curve, we need to calculate the cumulative frequencies. The cumulative frequency of a particular score, x is the number of scores less than or equal to x. In calculating the cumulative frequencies, we use the Upper-Class Boundaries as the highest possible score in an interval.

In the table below, the number of scores that is less than 10.5 is 0, the number less than 15.5 is 0+2 = 2, 20.5 is 2+5 = 7 and so on. So, the cumulative frequency of 20.5 is 7. By adding the frequencies, in turn, we obtain their cumulative frequencies.

Marks	Upper class boundary(x)	Frequency	Cumulative frequency
6 – 10	10.5	0	0
11 – 15	15.5	2	2
16 – 20	20.5	5	7
21 – 25	25.5	10	17
26 – 30	30.5	14	31
31 – 35	35.5	9	40
36 – 40	40.5	5	45
41 – 45	45.5	3	48
46 – 50	55.5	2	50
		50	

To draw a cumulative frequency curve, we use the following steps:
1. Determine the Upper-Class Boundaries for each interval.
2. Calculate the cumulative frequencies.
3. Plot the cumulative frequency against the upper-class boundary of each interval.
4. Join the points with a smooth curve.

The cumulative frequency curve for the data is shown below. The points plotted are (10.5, 0), (15.5, 2), (20.5, 7), (25.5, 17),….

The median is determined by drawing a horizontal line at 25 (half of 50) and a vertical line from the point where the line meets the curve. In this example, the median score is 27.5 (shown in the graph).

In a similar fashion, the upper and lower quartiles can be determined by drawing appropriate lines at 37.5 (three-quarters of 50) and 12.5 (one-quarter of 50) respectively. In this example, the lower quartile is estimated as 23.5, and upper quartile as 32.5.
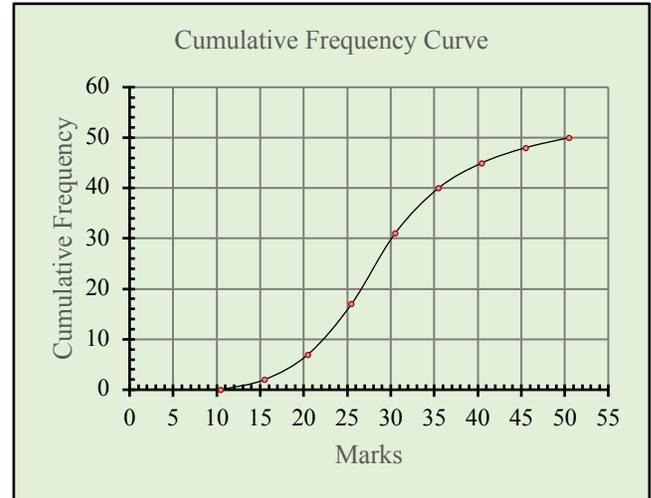
Interpretation of the median and the quartiles

In the above example, a median score of 27.5 indicates that one half the number of students in the class scored below 27.5 on the test. Of course, it also means that half of them scored more than 27.5 marks.

The lower quartile of 23.5 means that one quarter of the students scored below 23.5 on the test.

The upper quartile of 32.5 means that three-quarters of the students scored below 32.5 on the test.

In the same way, we can use this curve to find out:
a) The number of students who scored below 18 marks on the test?
b) The pass mark if 8 students qualify for the next round?

To answer part (a), we draw a vertical line starting at 18 marks to meet the curve and a horizontal line to meet the y-axis. The frequency of 6 indicates that 6 persons scored below 18 marks on the quiz.

To answer part (b), we must think of the number of students who will not qualify for the next round. This is calculated as 50 – 8 = 42 students. We must then draw a horizontal line at 42 from the y-axis until it meets the curve, then draw a vertical line from this point to meet the x-axis. The line will cut the axis at 36.5 marks. Hence, students scoring over 36.5 marks will qualify for the next round.

The standard deviation

The most accurate measure of spread is the standard deviation. It is widely used in statistics when variables are in interval or ratio scales. Intuitively, the standard deviation is an index which indicates how far the scores in the data set are from the mean. The difference between a score and the mean score is called the deviation.

In computing the standard deviation, we obtain all the deviations from the mean and calculate their average. A large standard deviation indicates that the data is spread out and there is a lot of variability in the scores. A small standard deviation indicates that the scores are close or homogenous.

The computation of the standard deviation is not as straightforward as it may seem because the deviations from the mean can be positive or negative. To calculate the average deviation, we square the deviations to remove their negative signs, find the average of the squared deviations and then find the square root so that the units are consistent.

A simple example will illustrate these procedures. Consider the set of raw scores, denoted x_i, obtained from marks on a quiz out of 10.

x_i = 4, 0, 6, 8, 5, 7. $n = 6$

1. Calculate the mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4+0+6+8+5+7}{6} = 5$$

2. Calculate the deviations from the mean, $x - \bar{x}$. See column 2 in the table below.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
0	$(0-5)$	25
4	$(4-5)$	1
5	$(5-5)$	0
6	$(6-5)$	1
7	$(7-5)$	4
8	$(8-5)$	9
		$\sum(x_i - \bar{x})^2 = 40$

3. Square the deviations and find their sum. See column 3.

4. Compute the standard deviation, s^2

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{40}{6} = 6.7$$

Hence, $S = \sqrt{\dfrac{40}{6}} = 2.6$

Interpreting the standard deviation

Researchers can tell a lot about how scores are distributed in a group by knowing the mean and standard deviation. In any population, most characteristics follow a simple pattern. Consider the variable, height of an adult in any population. The majority of adults are around average height, tall and short persons would be fewer in number, but very tall and very short persons are even fewer in number. In statistics, the actual percentages are known and this is called the **empirical rule**.

In any large population, almost all characteristics are normally distributed. When we know the mean and standard deviation of a variable in a population, we can apply the empirical rule which states with a high level of confidence that approximately

1. 68% of the population has scores within one standard deviation of the mean.
2. 95%percent of the population has scores within two standard deviations of the mean.
3. 99.7of the population has scores that fall within three standard deviations of the mean.
4. 0.3% of the population has scores that do not fall within three standard deviations from the mean.

Let us apply the empirical rule using actual statistics on the heights of male adults in the US.

Data
Mean height of adult males is 177 cm
Standard deviation for adult males is 7 cm

Interpretation

1. 68% of the adult males in the US have a height within the range 177 ± 7 cm, that is between 170 and 184 cm.
2. 95% of the adult males in the US have heights within the range $177 \pm 2(7)$cm, that is between 163 and 191 cm.
3. 99.7 % of adult males in the US have heights within the range $177 \pm 3(7)$cm, that is between 156 and 198 cm.
4. 0.3 % of the population has heights that are either more than 198 cm (177+21) or o heights that are less than 156 cm (177− 21).

We can represent this data using a diagram which shows how the entire population is divided using the empirical rule. The bell-shaped curve below encloses an area that represents an entire population. The vertical lines divide the area into the regions whose percentages are shown.
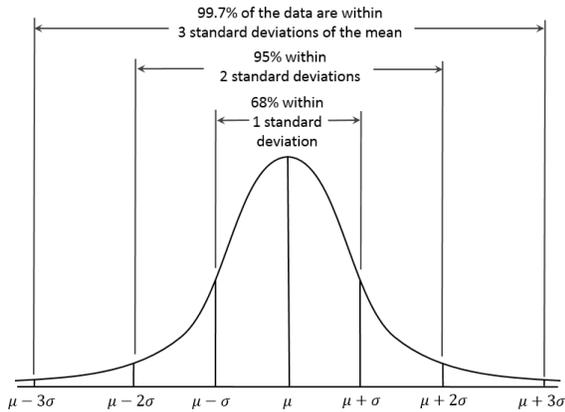
In the diagram, μ represents the population mean and σ represents the population standard deviation.

In the diagram below, the percentages in each of the six regions are shown. The percentages are given correct to 2 decimal places. The curve is symmetrical so the 68% that lies within one standard deviation of the mean is divided so that half – 34% lies one standard deviation above the mean and 34% lies one standard deviation below the mean.

Similarly, one half of 95% is 47.5%. We see on one side of the curve the two percentages add up to 47.5%, that is 34%+13.5% = 47.5%. This means that 47.5% of the population lies two standard deviations above the mean and 47.5% lies two standard deviations below the mean.
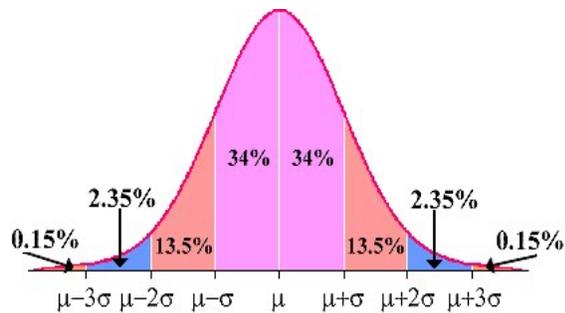
Adding 34 + 13.5 +2.35 = 49.85. This is one half of 99.7% and so, 49.85% of the population has scores that are 3 standard deviations above the mean and 49.85% have scores that are 3 standard deviations below the mean. A very small percentage (0.15 on

both sides) has scores that are more than 3 standard deviations from the mean.

Using the standard deviation to compare data sets

We can use the mean and standard deviation to compare scores belonging to two or more data sets. Consider the following marks made by Emma on four tests. We will compare Emma's performance to determine her relative standing with respect to the class.

|	Emma's Mark	Mean mark	Standard Deviation
English	55	65	10
Mathematics	59	51	4
Science	65	65	4

English
Emma scored 55 in English and the mean mark was 65. She scored 10 marks below the mean. In English, one standard deviation is 10 marks, so Emma's mark was 1 standard deviation below the mean. This means that 84% of the class did better than she or she was better than only 16% of the class in English.
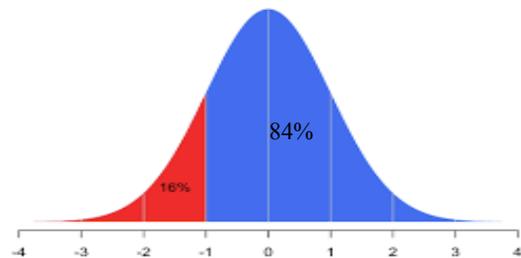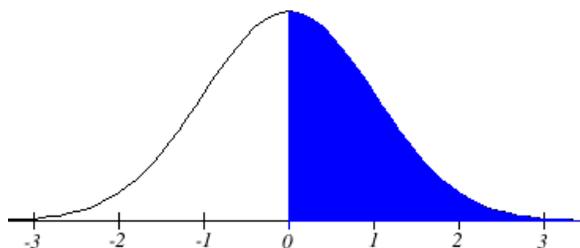

Mathematics
Emma scored 59 in Mathematics and the mean mark was 51. She scored 8 marks above the mean. In Mathematics, one standard deviation is 4 marks, so Emma's mark was 2 standard deviations above the mean. This means that only 2.5% of the class did better than she or she was better than 97.5 % of the class in Mathematics.
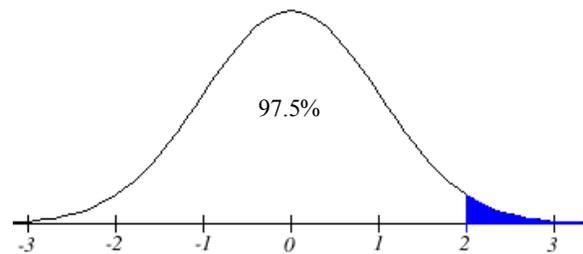
Science

Emma scored 65 in Science and the mean mark was 65. Her score was actually the same as the mean score. This means that exactly 50% of the class did better than her.

The above analysis indicates that relative to her class, Emma's performance in Mathematics was best. Her mathematics score was better than 99.7% of her class. Her performance in Science was also better than her performance in English because her score in Science was better than 50% of the class while her score in English was better than only 16% of her class.

PROBABILITY

We often encounter situations in which we need to take a chance or risk and we are faced with a set of options. The study of probability gives us the tools to make informed decisions about areas in life, where there is a degree of uncertainty. It teaches us to consider all possible conditions and the relative importance of each one to our final outcome. We begin by learning how to calculate the probability of events that occur in simple experiments.

The probability scale

Probability is the measure of how likely an event is to occur. In mathematics, we quantify measures by assigning some value to it. In constructing a probability scale, we can think of the two extremes

- What is the probability of an event that will never happen? This is an impossible event. It is assigned the smallest possible value of zero.

- What is the probability of an event that is certain to happen? This is the certain event. It is assigned the largest possible value of one.

Hence, for any event, the probability will always lie between 0 and 1. No value of probability can ever be negative or greater than one.

We can think of the probability scale as a number-line, with values only in the interval between 0 and 1.
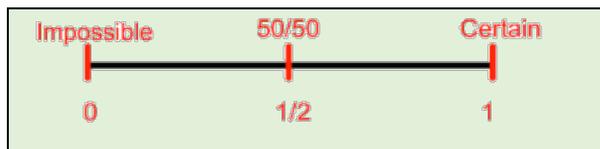
In the middle of the scale, the probability is one-half and will correspond to events that have two possible outcomes, either of which is equally likely to occur. Examples of such events are getting a Head when flipping a coin or choosing a red card from a deck of playing cards. These events are said to have a 50% probability of occurring, and so there is a 50-50 or even chance for either outcome to happen. [Notice that probability can also be represented by a percentage.]

Terms in probability

The following are definitions of common terms used in probability.

An **experiment** is a situation involving chance or probability.

The possible results of an experiment are called **outcomes**.

An **impossible event** is one that will never occur.
Example: A rose bush growing from an apple seed is an impossible event.

A **certain event is one** that will definitely occur.
Example: The sun rising every morning is a certain event.

An event whose probability is close to one is called a **very likely event**
Example: The probability that a teenager will own a cell phone is very likely.

Unlikely Event
An event whose probability is close to zero.
Example: Buying a winning ticket in the lotto.

Measuring Probability

To measure the likelihood that an event will happen in simple experiments, we consider the conditions that affect the outcomes and use a fraction to compare the number of favourable events to the total number of events.

If E is an event, then the probability of E occurring is written as P(E). This is read as P of E.

P(E) is defined as:

$$\frac{\text{The number of outcomes that are favourable to the event}}{\text{The number of possible outcomes}}$$

OR

$$\frac{\text{The number of ways that the event can take place}}{\text{The number of possible outcomes}}$$

Let us consider how this definition applies to calculating the probability of an event. The diagram below shows a spinner with 4 equal sized sectors, differently coloured as yellow, blue, green and red. An experiment of turning the spinner will result in the arrow pointing in one of the regions – blue, red, orange or green. Since all regions are equal in area, all four outcomes are equally likely and so the probability of the spinner landing on any colour is the same.
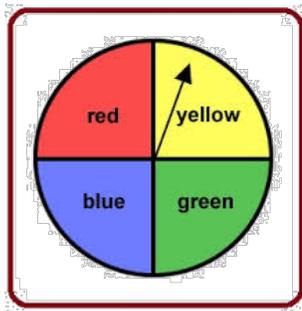
In this example, the *experiment* is, spinning the spinner.

An *outcome* is the result of a single *trial* of an experiment. A trial will comprise one turn of the spinner.

The *sample space* is the set of all possible *outcomes*. This experiment has a sample space of 4 outcomes {red, blue, green, yellow}.

An *event* consists of one or more outcomes of an experiment. For example, one event of this experiment is landing on blue. Another event is landing on blue or red.

We know that the total number of possible outcomes is four (yellow, blue, green or red). So, we can now calculate the following probabilities:
   a) What is the probability of landing on blue?
   b) What is the probability of landing on blue or red?

   a) The probability of the spinner landing on blue is, $\frac{1}{4}$. The numerator takes the value of 1 since there is only one outcome that is favourable (one sector is blue). The denominator takes the value of 4 since there are 4 possible outcomes.

   b) The probability of the spinner landing on blue or red is $\frac{2}{4}$ or $\frac{1}{2}$ Here, the numerator takes the value of 2 since there are two favourable outcomes (two sectors are possible, blue or red). Just as before, the denominator takes the value of 4 since there are 4 possible outcomes.

Equiprobable or equally likely events

In tossing of a fair coin, the probability of getting heads is the same as the probability of getting tails. In the same way, in rolling a fair die, the probability of getting a 6 is the same as the probability of getting a 1, 2, 3, 4 or 5. Such events are called **equiprobable** events or equally likely events.

If, however, the die or the coin is weighted (sometimes called loaded), the outcomes will not be equally likely. In such a case, we refer to the coin or die as **biased**. Otherwise, the coin or die is said to be ordinary or fair or **unbiased**.

Three common experiments involving equiprobable events are described below.

A fair coin is tossed**.**
There are two possible outcomes, Heads (H) and tails (T). The probability of obtaining a head is P(H) and the probability of obtaining tails is P(T).

$$P(H) = \frac{1}{2} \qquad P(T) = \frac{1}{2}$$

In each case, there is one favourable outcome out of 2 possible outcomes- Head and Tail.

A fair die is tossed. Since there are six equiprobable events, we can list the probability of each possible outcome as follows:

$$P(1) = \frac{1}{6} \qquad P(2) = \frac{1}{6} \qquad P(3) = \frac{1}{6}$$
$$P(4) = \frac{1}{6} \qquad P(5) = \frac{1}{6} \qquad P(6) = \frac{1}{6}$$

If A is the event that an even number is obtained, then $P(A) = \frac{3}{6}$ as there are three even numbers out of 6 possible numbers.

A card is picked from an ordinary pack of playing cards. Q is the event that a Queen is obtained.

There are four Queens in the pack and fifty-two cards in all.

Hence, $P(Q) = \dfrac{4}{52}$, which may be reduced to $\dfrac{1}{13}$.

Impossible and certain events

An ordinary fair die is tossed. Find the probability that a '7' occurs

$$P(7) = \frac{\text{The number of sevens(7) on a die}}{\text{The number of possible outcomes}}$$
$$= \frac{0}{6}$$
$$= 0$$

We can think of other impossible events. Say, for example, a card is chosen at random from an ordinary pack of playing cards. The probability that the card is 'Green' is impossible, since all of the 52 cards in an ordinary pack of cards are either red or black.

An ordinary, fair die is tossed. Find the probability of obtaining a whole number.

P(Whole number)
$$= \frac{\text{The number of whole numbers on a die}}{\text{The number of possible outcomes}} = \frac{6}{6} = 1$$

This event must occur, is sure to occur or is certain to occur, since we shall always get a whole number, (1,2,3,4,5 and 6 are all whole numbers).

A month of the year is chosen at random. Find the probability that it has more than 27 days.

P(month has >27 days)
$$= \frac{\text{The number of months with >27 days}}{\text{The number of months}} = \frac{12}{12} = 1$$

We can think of other sure or certain events. Say, for example, a student is chosen at random from an 'all boys school'. The probability that the student is a boy is certain since all the students are boys. Both the numerator as well as the denominator shall have the same value and the fraction clearly reduces to 1.

Laws of probability

We are now in a position to formally state the laws of probability.

1. The probability of an event, $P(A)$ must lie in the interval $0 \le P(A) \le 1$.
2. The sum of the probabilities of all the outcomes in a sample space must total one. $P(S) = 1$.
3. If two events, A and A' make up a sample space, then $P(A') = 1 - P(A)$.

Example 12

A card is drawn at random from an ordinary deck of 52 cards.
S – the event an Ace is obtained
Calculate the probability of not obtaining of an Ace,

Solution

P(S').
$$P(S) = \frac{4}{52} = \frac{1}{13} \qquad \text{Hence,} \quad P(S') = 1 - \frac{1}{13} = \frac{12}{13}$$

Example 13

The probability, $P(R)$, of rain tomorrow is given by
$$P(R) = \frac{5}{7}.$$
Calculate $P(R')$, the probability that it will not rain tomorrow.

Solution

$$P(R') = 1 - P(R) = 1 - \frac{5}{7} = \frac{2}{7}$$

Experimental and theoretical probability

The theoretical probability of an event is a predicted probability. The value is based on the ratio of the number of outcomes that are favourable to the total possible outcomes (denominator) in the sample space.

P(E)
$$= \frac{\text{The number of outcomes that are favourable to an event}}{\text{The number of possible outcomes}}$$

Experimental probability, on the other hand, is the value of probability that was obtained from the result of an experiment. It is based on what actually occurred after a number of trials.

In an experiment, a die was tossed 120 times. The results are shown in the table below.

Outcome (E)	1	2	3	4	5	6
Frequency	24	18	19	21	16	22

For each outcome, we can calculate the experimental probability, by using the following ratio.

$$P(E) = \frac{\text{Frequency of the event}}{\text{Total number of trials}}$$

This ratio gives the experimental probability. It is different for each outcome because the frequencies are different.

Using theoretical probability, the probability of getting a six or any other number when tossing a die is the same because each number occurs once and the number of possibilities is always 6. It is calculated as:

$$P(6) = \frac{1}{6} = 0.167.$$

Outcome	Frequency	Experimental Probability	Theoretical Probability
1	24	0.2	0.167
2	18	0.15	0.167
3	19	0.158	0.167
4	21	0.175	0.167
5	16	0.133	0.167
6	22	0.183	0.167

Notice that the experimental probability value approximates in value to the theoretical probability of 0.167, but not equal to it. In fact, if we toss the die 600 times, the experimental probability value will be even closer to the theoretical probability value. As the number of trials increases, the closer the experimental probability approaches its theoretical value. In some situations, we have no choice but to use experimental probability in predicting events.

Example 14

A company inspector found that 3 in every 100 bulbs manufactured are defective.
  i.   What is the probability that a bulb is defective?
  ii. How many bulbs are expected to be defective in a shipment of 2000 bulbs?

Solution

  i.   The probability that a bulb is defective
  $$= \frac{\text{The number of defective bulbs}}{\text{The number of bulbs in the sample}} = \frac{3}{100}$$

  ii. Number of bulbs expected to be defective in a shipment of 2000 bulbs
  $$= \frac{3}{100} \times 2000 = 60$$

Example 15

In an experiment, an ordinary, fair die is tossed a certain number of times. The results are shown in the table below.

Result	Frequency
1	3
2	6
3	4
4	6
5	4
6	7

  i.   Based on this data, what is the experimental probability of tossing a 4?
  ii. What is the theoretical probability of tossing a 4?

Solution

  i.   Frequency of 4 = 6
  Total number of trials = 3 + 6 + 4 + 6 + 4 + 7 = 30
  $$\text{The probability of obtaining a 4} = \frac{6}{30} = \frac{1}{5}$$

  ii. Theoretical probability of obtaining a 4 = $\frac{1}{6}$.

Mutually Exclusive Events

Two events are mutually exclusive if they cannot occur together. This means that they have no common outcomes or they are disjoint. Some examples of mutually exclusive events are:

1. Tossing a coin – the outcomes Head and Tail cannot occur together.
2. Tossing three coins – the eight outcomes in the sample space are mutually exclusive.
    {HHH, TTT, HTH, HTT, HHT, THT, THH, TTH}
    Only one of these is possible in a single toss.

3. Tossing a die – the six outcomes $\{1, 2, 3, 4, 5, 6\}$ in the sample space are mutually exclusive and no two can occur simultaneously.
4. Selecting an Ace or a Jack from a deck of playing cards – these two events are mutually exclusive since no card can be an Ace and a Jack. However, the event - selecting an Ace and a black card is not mutually exclusive because these events have common outcomes. The set of Aces contain two black cards.
5. In rolling two dice, the events 'the sum is 4' and the event 'at least one is 5' are mutually exclusive because there are no outcomes common to both events. The diagram below shows all 36 outcomes with the event 'the sum is 4' shaded red and the event 'at least one is 5' shaded green. Notice that the sets are disjoint.

Second Die

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

First Die

Addition Law for mutually exclusive events

Consider the experiment – tossing a die. We wish to determine the probability of getting a six or a two. This can be done by simply applying the definition of probability. There are 2 favourable outcomes out of a total of 6 outcomes, so that:
$$P(6 \text{ or } 2) = \frac{2}{6}$$
Alternatively, we could have obtained the same answer by combining the probabilities of the two events as shown below.
$$P(6) = \frac{1}{6} \qquad P(2) = \frac{1}{6}$$
$$P(6 \text{ or } 2) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

Consider the experiment tossing 3 coins. We wish to determine the probability of getting either 3 heads or 3 tails. The sample space is shown below:

$$\{\text{HHH}, \text{TTT}, \text{HTH}, \text{HTT}, \text{HHT}, \text{THT}, \text{THH}, \text{TTH}\}$$

There are 2 favourable outcomes out of 8 outcomes.
$$P(3H \text{ or } 3T) = \frac{2}{8}$$

Alternatively, we could have obtained the same answer by combining the probabilities of the two events as shown below.

$$P(3H) = \frac{1}{8} \qquad P(3T) = \frac{1}{8}$$
$$P(3H \text{ or } 3T) = \frac{1}{8} + \frac{1}{8} = \frac{2}{8}$$

In the above examples, we were calculating the combined probability of two mutually exclusive events using two methods. We saw that addition rule is more efficient because it did not require us to list the entire sample space. In practice, we do not always have knowledge of the sample space and so the addition rule is very useful in calculating the probability of mutually exclusive events.

Addition law for mutually exclusive events
$$P(E \text{ or } F) = P(E) + P(F)$$
$$P(E \cap F) = 0$$

E

F
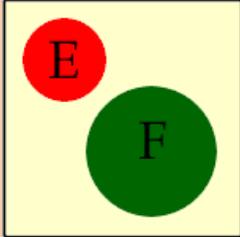
Example 16

It is known that the probability of obtaining zero defectives in a sample is 0.34 whilst the probability of obtaining 1 defective item in the sample is 0.46. What is the probability of obtaining either zero or one defective item in the sample?

Solution

Let event E_1 be "obtaining zero defectives" and E_2 be "obtaining 1 defective item".
$$P(E_1) = 0.34$$
$$P(E_2) = 0.46$$
Events E_1 and E_2 are mutually exclusive, so
$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) = 0.34 + 0.46 = 0.8$

Addition Law for non-mutually exclusive events

The addition rule for mutually exclusive events stated above is a special case of the general addition rule in probability. The general addition rule applies to both mutually exclusive events as well as non-mutually exclusive events. It is related to the following law in set theory when we have two intersecting sets.

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

If we divide each term by the total number of outcomes in the sample space, we obtain
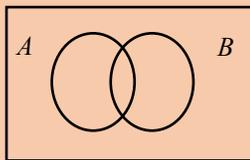
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

For mutually exclusive events, $P(A \cap B) = 0$, In this respect, this law holds for both type of events.

Addition Rule for non-mutually exclusive events
When two events, A and B, are non-mutually exclusive, the probability that A or B will occur is:
$$P(A \ or \ B) = P(A) + P(B) - P(A \ and \ B)$$

Example 17

A single card is chosen at random from a standard deck of 52 playing cards. What is the probability of choosing a king or a club?

Solution

$P(\text{King}) = \frac{4}{52}$ $P(\text{Club}) = \frac{13}{52}$

These are not mutually exclusive events, so we use the general addition rule.

P(King or Club)
$= P(King) + P(Club) - P(King \ and \ Club)$
$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$
$= \frac{16}{52} = \frac{4}{13}$

Example 18

In a math class of 30 students, 17 are boys and 13 are girls. On a test, 4 boys and 5 girls made an A grade. If a student is chosen at random from the class, what is the probability of choosing a girl or an A student?

Solution

$P(\text{girl}) = \frac{13}{30}$ $P(A) = \frac{9}{30}$ $P(\text{girl and A}) = \frac{5}{30}$

$P(\text{girl or A}) = P(\text{girl}) + P(A) - P(\text{girl and A})$
$= \frac{13}{30} + \frac{9}{30} - \frac{5}{30} = \frac{17}{30}$

Independent Events

Two events are said to be independent of each other if the probability of one event occurs is not affected by the probability of the other event occurring. In tossing two coins, the event that a head occurs on the first coin and a tail on the second coin are independent events. Notice the outcome of the first coin (getting a Head) does not influence the outcome of the second coin (getting a Tail).
Some examples of independent events:
1. If we select a card from a deck of playing cards and then roll a die, then these are independent events. The card selected does not influence the outcome of the die.
2. The event that the first child in a family is a boy and the second child is also a boy. The gender of the first child does not influence the gender of the second child.
3. The event that it rains today and the event that today is a Sunday. Rain can fall on any day of the week and these events do not influence each other.
4. If we select a marble from a jar of 30 coloured marbles (10R, 10B, 10Y), replace it and then select another one, then these are independent events. Since the first marble is replaced the jar is much the same for the second selection. The outcome of the second selection is not affected by what was obtained on the first. If, however, the first marble was not replaced then the composition of the jar changes and this will affect the outcome of the second selection. The events will no longer be independent, but dependent.

Multiplication law- independent events

Consider an experiment where a coin is flipped and a die is tossed.
Let H represent the event that a head is obtained when the coin is flipped and
Let S represent the event that a six (6) is obtained when the die is tossed.
We wish to calculate the probability that both a head and a six occurs.

143

We can approach this task using our prior knowledge of basic probability. Our sample space has the following outcomes:

{(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6)}

The number of favourable outcomes is one and there are 12 outcomes in all. Hence,

$$P(H \text{ and } 6) = \frac{1}{12}.$$

Now, we can arrive at the same answer by combining the two probabilities as follows:

$$P(H) = \frac{1}{2}$$
$$P(6) = \frac{1}{6}$$
$$P(H \text{ and } 6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$$

Let us look at another example.

You roll a fair die twice. What is the probability of rolling a three on the first roll and an even number on the second roll?

Using the sample space for two dice, we note that there are 36 outcomes.

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

To get a five on the first roll and an even number on the second roll, we have the following three favourable outcomes:

$$(5, 2), (5, 4), (5, 6)$$
$$P(5 \text{ and Even No}) = \frac{3}{36}$$

Now, we can arrive at the same answer by combining the two probabilities as follows:

$$P(5) = \frac{1}{6}$$
$$P(\text{Even No}) = \frac{3}{6}$$
$$P(5 \text{ and Even No}) = \frac{1}{6} \times \frac{3}{6} = \frac{3}{36}$$

The above examples illustrate that we can obtain the combined probability of two independent events by multiplying their probabilities. This method is not only more efficient but particularly useful when we do not have knowledge of the sample space.

We are now in a position to formally state the multiplication law of probability.

Multiplication law of probability- independent events
If A and B are two independent events, then
$P(A \text{ and } B) = P(A) \times P(B)$.

This rule can be extended to two, three or any number of independent events. For example,
If A, B and C are all independent events, then
$$P(A \text{ and } B \text{ and } C) = P(A) \times P(B) \times P(C)$$

Example 19
Suppose that we draw a card from a standard deck, replace this card, shuffle the deck and then draw again. What is the probability that both cards are kings?

Solution
The probability of drawing a king for the first card is $\frac{1}{13}$.
The probability of drawing a king on the second draw is $\frac{1}{13}$. Since the events are independent,
$$P(King \text{ and } King) = \frac{1}{13} \times \frac{1}{13} = \frac{1}{169}$$

Example 20
Three events are defined as follows:
H – a head is obtained when a coin is tossed.
F – a five (5) is obtained when a die is tossed.
A – an ace is obtained when a card is picked from a pack.
Calculate the probability that all three events occur.

Solution
These are independent events. Let
P(a head and a five and an ace)
$= P(H \text{ and } F \text{ and } A) = P(H \cap F \cap A)$.
$$P(H \cap F \cap A) = P(H) \times P(F) \times P(A)$$
$$= \frac{1}{2} \times \frac{1}{6} \times \frac{4}{52} = \frac{1}{156}$$