# Mathematics

*Senior 2 Part II*

MELVIN CHIA

Started on 1 January 2023

Finished on ...

# Contents

# Chapter 18

# Statistics

## 18.1 Basic Concepts

Statistics mainly study how to collect, organize, summarize, and interpret data. It is a branch of mathematics that deals with the collection, analysis, interpretation, and presentation of data. It is used to answer questions about the data and to make decisions based on the data.

### Population and Sample

In statistics, a population is the entire group of individuals that we are studying, and the units that form a population are called individuals or elements. A sample is a subset of the population. The number of elements in a sample is called the sample size. For example: select 20 of the 4,000 senior high school mathematics UEC exam papers and record their scores:

| | | | | |
|---|---|---|---|---|
| 72 | 80 | 96 | 20 | 42 |
| 75 | 60 | 92 | 18 | 53 |
| 82 | 77 | 53 | 29 | 34 |
| 57 | 79 | 82 | 90 | 41 |

Here, the population is the 4,000 scores, each of which is an element of the population. The sample is the 20 scores, the sample size is 20.

### Census and Sample Survey

The way of surveying can be divided into two types: census and sample survey. A census is a survey in which every element of the population is included in the sample. For example: national census. The data collected in a census is more accurate and reliable, but it is very expensive and time-consuming.

A sample survey is a survey in which only a part of the population is included in the sample. Researchers can use a sample survey to estimate the characteristics of the population. For example: a light bulb manufacturer produces a lot of light bulbs, thus it is impossible to test every single light bulb. The manufacturer can randomly select a sample of light bulbs and test them.

## 18.2 Data Processing

Data that are collected must be processed before they can be analyzed.

### Frequency Distribution

When the possible values of a dataset are not too many, we can use a frequency distribution table to organize the data. The frequency distribution table is a table that shows the frequency of each value in a dataset. The frequency of a value is the number of times that value appears in the dataset.

When there are too many possible values, we must group the values into classes. Before grouping the values, we must first determine the range of the values, aka the difference between the largest and smallest values, then determine the number of classes. The number of classes should be determined according to the purpose of the study and the identity of the data. After classifying the data, the range of each group is called the class interval. Typically, the class interval is the same for all classes, and must be greater than the number of classes divided by the range of the data. After the number and interval of the classes are determined, we can arrange the frequency of each class in a frequency distribution table.

Take 100 sample from a population of some kind of component, their weight (in $g$), are as below:

| | | | | | |
|---|---|---|---|---|---|
| 1.36 | 1.49 | 1.43 | 1.41 | 1.37 | 1.40 |
| 1.32 | 1.42 | 1.47 | 1.39 | 1.41 | 1.36 |
| 1.40 | 1.34 | 1.42 | 1.42 | 1.45 | 1.35 |
| 1.42 | 1.39 | 1.44 | 1.42 | 1.39 | 1.42 |
| 1.42 | 1.30 | 1.34 | 1.42 | 1.37 | 1.36 |
| 1.37 | 1.34 | 1.37 | 1.37 | 1.44 | 1.45 |
| 1.32 | 1.48 | 1.40 | 1.45 | 1.39 | 1.46 |
| 1.39 | 1.53 | 1.36 | 1.48 | 1.40 | 1.39 |
| 1.38 | 1.40 | 1.36 | 1.45 | 1.50 | 1.43 |
| 1.38 | 1.43 | 1.41 | 1.48 | 1.39 | 1.45 |

| | | | | | |
|---|---|---|---|---|---|
| 1.37 | 1.37 | 1.39 | 1.45 | 1.31 | 1.41 |
| 1.44 | 1.44 | 1.42 | 1.47 | 1.35 | 1.36 |
| 1.39 | 1.40 | 1.38 | 1.35 | 1.38 | 1.43 |
| 1.42 | 1.42 | 1.42 | 1.40 | 1.41 | 1.37 |
| 1.46 | 1.36 | 1.37 | 1.27 | 1.37 | 1.38 |
| 1.42 | 1.34 | 1.43 | 1.42 | 1.41 | 1.41 |
| 1.44 | 1.48 | 1.55 | 1.39 | | |

In the dataset above, the minimum value is 1.27 and the maximum value is 1.55.

$\therefore$ The range of the data is $1.55 - 1.27 = 0.28$.

If we classify the data into 10 classes, then the class interval must be greater than $\frac{0.28}{10} = 0.028$. Thus, we can use a class interval of 0.03.

Let the lower limit of the first class be 1.27, then the lower limit of the second class is $1.27 + 0.03 = 1.30$.

Since all the values in the dataset are of 2 decimal places, the upper limit of the first class is should be 1.29. By the same logic, we can get all the classes: $1.27 - 1.29$, $1.30 - 1.32$, $\cdots$, $1.54 - 1.56$.

Now we can arrange the data into the frequency distribution table:

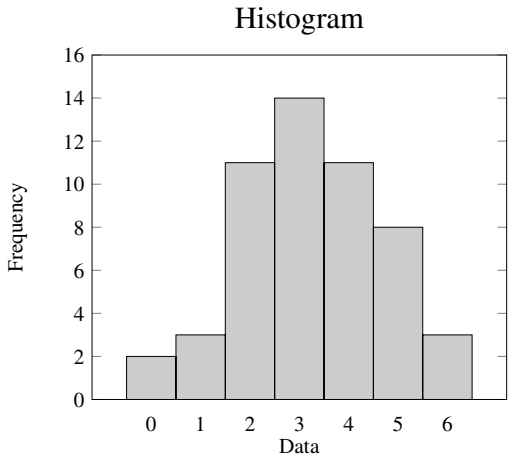
| Weight $m(g)$ | Frequency |
|---|---|
| $1.27 - 1.29$ | 1 |
| $1.30 - 1.32$ | 4 |
| $1.33 - 1.35$ | 7 |
| $1.36 - 1.38$ | 22 |
| $1.39 - 1.41$ | 24 |
| $1.42 - 1.44$ | 24 |
| $1.45 - 1.47$ | 10 |
| $1.48 - 1.50$ | 6 |
| $1.51 - 1.53$ | 1 |
| $1.54 - 1.56$ | 1 |

In the example above, we assume that the weight of the components is accurate to 2 decimal places. Hence, if a component has a weight of $1.443g$, it is rounded to $1.44g$, thus it belongs to the class $1.42 - 1.44$. Hence, the actual range of the first class $1.27 - 1.29$ is $1.265 \leq m < 1.295$, written as $1.265 - 1.295$, while 1.265 and 1.295 are the boundaries of the first class, 1.265 is the lower boundary and 1.295 is the upper boundary. The mean of the lower boundary and upper boundary of a class is called the class midpoint. For example, the class midpoint of the first class is $\frac{1.265+1.295}{2} = 1.28$.

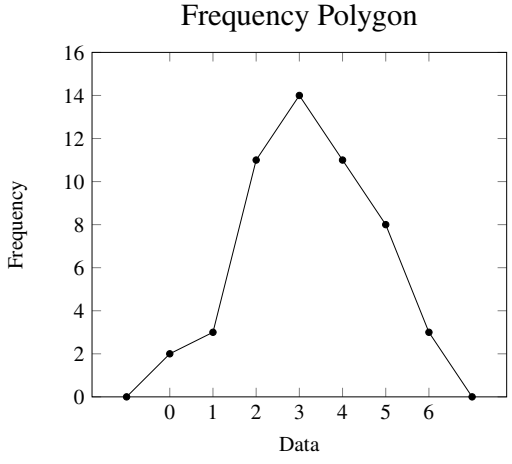
When we are analyzing the data data that have been classified into classes, the midpoint of each class is used as the representative value of the class. Thus, we should try our best to make the data-intensive place the group midpoint when choosing the class interval and boundaries, so that the data can be analyzed more precisely.

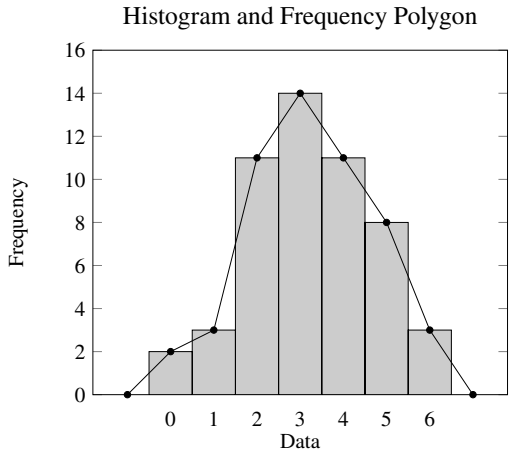
The distribution of frequency can be represented by a histogram or a frequency polygon.

The histogram is a row of continuous bars, the bottom side of each bar on the x-axis. For unclassified data, the bottom side of each bar is marked with the values, while the height of each bar is the frequency of the corresponding value. For classified data, the bottom side of each bar is marked with the boundaries of the corresponding class, while the area of each bar must be proportional to the frequency of the corresponding class. When the class interval of each class is the same, we can use the frequency of each class as the height of the bar.




The frequency polygon is a continuous line graph, the x-axis is the midpoint of each class, and the y-axis is the frequency of each class. To draw a frequency polygon, we plot each point, including the point before the first class and the point after the last class that uses 0 as their frequency, and then connect the points with a continuous line.






### 18.2.1 Practice 1

There are 105 students in a senior 3 art and commerce class. In a mock exam of UEC, their scores for Mathematics subject are as follows:

| | | | | | | |
|---|---|---|---|---|---|---|
| 35 | 88 | 67 | 32 | 38 | 34 | 45 |
| 78 | 54 | 58 | 69 | 21 | 90 | 78 |
| 74 | 43 | 42 | 35 | 57 | 34 | 77 |
| 89 | 66 | 74 | 71 | 44 | 56 | 48 |
| 33 | 24 | 73 | 63 | 51 | 59 | 49 |
| 34 | 55 | 52 | 75 | 72 | 62 | 62 |
| 44 | 48 | 73 | 49 | 57 | 67 | 80 |
| 70 | 66 | 54 | 32 | 29 | 35 | 37 |
| 47 | 41 | 51 | 36 | 46 | 55 | 53 |
| 60 | 53 | 62 | 39 | 35 | 48 | 42 |
| 71 | 63 | 70 | 33 | 45 | 42 | 44 |
| 61 | 59 | 67 | 30 | 42 | 43 | 89 |
| 96 | 82 | 47 | 63 | 54 | 34 | 45 |
| 45 | 87 | 28 | 34 | 29 | 77 | 64 |
| 64 | 50 | 48 | 75 | 33 | 56 | 84 |

(a) Find the range of the data.

**Sol.**

$$\begin{aligned} \text{Max value} &= 96 \\ \text{Min value} &= 21 \\ \therefore \text{Range} &= 96 - 21 \\ &= 75 \end{aligned}$$

(b) Group the data into 10 classes, draw a frequency distribution table, and find the upper and lower boundary and midpoint of each class.

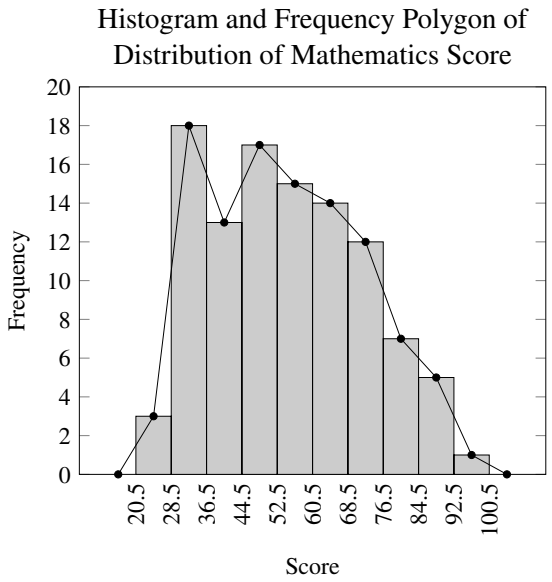
**Sol.**

$$\begin{aligned} \text{Range} &= 75 \\ \text{Number of classes} &= 10 \\ \text{Class width} &= \frac{75}{10} \\ &= 7.5 \\ &\approx 8 \end{aligned}$$

| Score | Lower | Upper | Mid | Freq. |
|---|---|---|---|---|
| 21 - 28 | 20.5 | 28.5 | 24.5 | 3 |
| 29 - 36 | 28.5 | 36.5 | 32.5 | 18 |
| 37 - 44 | 36.5 | 44.5 | 40.5 | 13 |
| 45 - 52 | 44.5 | 52.5 | 48.5 | 17 |
| 53 - 60 | 52.5 | 60.5 | 56.5 | 15 |
| 61 - 68 | 60.5 | 68.5 | 64.5 | 14 |
| 69 - 76 | 68.5 | 76.5 | 72.5 | 12 |
| 77 - 84 | 76.5 | 84.5 | 80.5 | 7 |
| 85 - 92 | 84.5 | 92.5 | 88.5 | 5 |
| 93 - 100 | 92.5 | 100.5 | 96.5 | 1 |

(c) Construct a histogram and frequency polygon.

**Sol.**




## Cumulative Frequency Distribution

Summing up the frequency of each class, we obtain the cumulative frequency distribution. Use the upper boundary of each class as the x-axis, and the cumulative frequency as the y-axis, we can draw the cumulative frequency distribution by plotting each point including the point before the first class that uses 0 as its frequency and connect them together. If we split the x-axis and the higest point of the curve into 100 equal

parts, we get the percentage of the cumulative frequency distribution.

### 18.2.2 Practice 2

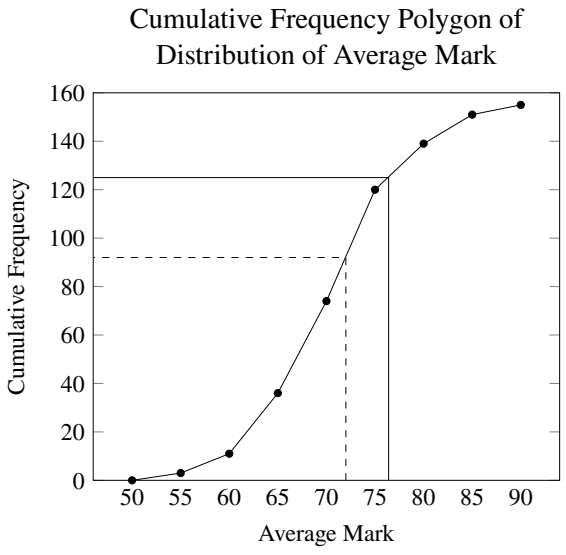
There are 155 students in a senior 3 art and commerce class, and the frequency distribution table of their average marks is shown below:

| Average Mark | Frequency |
|---|---|
| 50 - 55 | 3 |
| 55 - 60 | 8 |
| 60 - 65 | 25 |
| 65 - 70 | 38 |
| 70 - 75 | 46 |
| 75 - 80 | 19 |
| 80 - 85 | 12 |
| 85 - 90 | 4 |

(a) Make a cumulative frequency distribution table and draw a cumulative frequency polygon.

**Sol.**

| Avg | Freq. | Lower Than | Cumm. Freq. |
|---|---|---|---|
| 50 - 55 | 3 | 55 | 3 |
| 55 - 60 | 8 | 60 | 11 |
| 60 - 65 | 25 | 65 | 36 |
| 65 - 70 | 38 | 70 | 74 |
| 70 - 75 | 46 | 75 | 120 |
| 75 - 80 | 19 | 80 | 139 |
| 80 - 85 | 12 | 85 | 151 |
| 85 - 90 | 4 | 90 | 155 |




(b) If the average mark of a student is 72, find his rank in the class.

**Sol.**

In the graph above, we can see that there are approximately 92 students who have an average mark lower than 72. Therefore, the rank of the student is $155-92 = 63$.

(c) If the top 20% of the class are to be awarded a certificate, find the minimum average mark required for the certificate.

**Sol.**

$$\begin{aligned}\text{Top } 20\% &= 20\% \times 155 \\ &= 31\end{aligned}$$

Therefore, students with an average mark corresponding to cumulative frequency higher than 124 will be awarded a certificate.

In the graph above, The minimum average mark required for the certificate is 76.

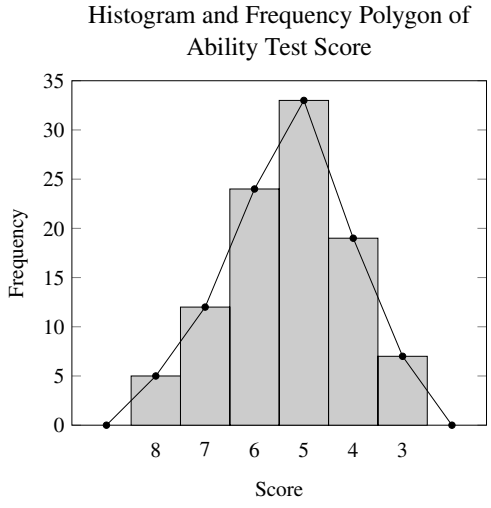
### 18.2.3 Exercise 18.2

1. A company performed an ability test on 100 job seekers and the results are shown in the following table:

| Score | 8 | 7 | 6 | 5 | 4 | 3 |
|---|---|---|---|---|---|---|
| Frequency | 5 | 12 | 24 | 33 | 19 | 7 |

Construct a hustogram and a frequency polygon for the data above.

**Sol.**




2. Take 120 ears of rice from a rice field, the length of each ear is measured (in $cm$) and the results are as fol-

lowing:

| | | | | | |
|---|---|---|---|---|---|
| 6.5 | 6.4 | 6.7 | 5.8 | 5.9 | 5.9 |
| 5.2 | 4.0 | 5.4 | 4.6 | 5.8 | 5.5 |
| 6.0 | 6.5 | 5.1 | 6.2 | 5.4 | 5.0 |
| 5.0 | 6.8 | 6.0 | 5.0 | 5.7 | 6.0 |
| 5.5 | 6.8 | 6.0 | 6.3 | 5.5 | 5.0 |
| 6.4 | 5.8 | 5.9 | 5.7 | 6.8 | 6.6 |
| 6.0 | 6.4 | 5.7 | 7.4 | 6.0 | 5.4 |
| 6.5 | 6.0 | 6.8 | 5.3 | 6.4 | 5.7 |
| 6.7 | 6.2 | 5.6 | 6.0 | 6.7 | 6.7 |
| 6.0 | 5.5 | 6.2 | 6.1 | 5.3 | 6.2 |
| 5.8 | 5.3 | 7.0 | 6.0 | 6.0 | 5.9 |
| 5.4 | 6.0 | 5.2 | 6.0 | 6.3 | 5.7 |
| 6.8 | 6.1 | 4.5 | 5.4 | 6.3 | 6.9 |
| 4.9 | 5.1 | 5.6 | 5.9 | 6.1 | 6.5 |
| 6.6 | 5.7 | 5.8 | 5.8 | 6.2 | 6.3 |
| 6.5 | 5.3 | 5.9 | 5.5 | 5.8 | 6.3 |
| 5.2 | 6.0 | 7.0 | 6.4 | 5.8 | 6.3 |
| 6.0 | 6.3 | 5.6 | 6.8 | 6.6 | 4.7 |
| 5.7 | 5.7 | 5.6 | 6.3 | 6.0 | 5.8 |
| 6.3 | 7.5 | 6.2 | 6.4 | 7.0 | 6.5 |

(a) Find the range of the dataset.

**Sol.**

$$\begin{aligned}\text{Min value} &= 4.0\\ \text{Max value} &= 7.5\\ \therefore \text{Range} &= 7.5 - 4.0\\ &= 3.5\end{aligned}$$

(b) Group the data into 12 classes, make a frequency distribution table, find the upper and lower boundaries and midpoint of each class, and calculate the cumulative frequency.

**Sol.**

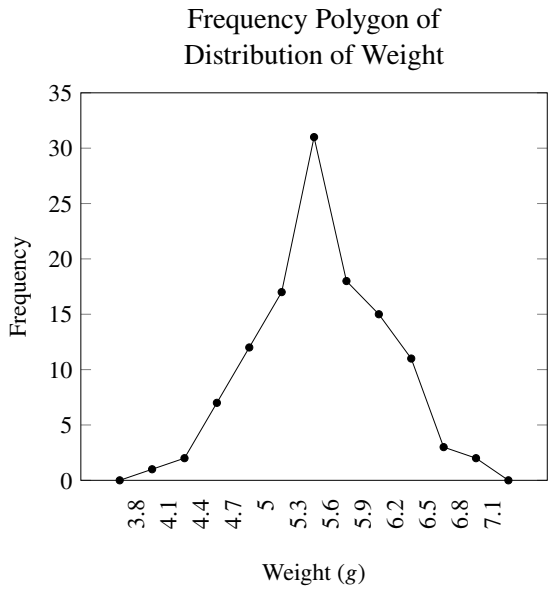
$$\begin{aligned}\text{Range} &= 3.5\\ \text{Number of classes} &= 12\\ \therefore \text{Class width} &= \frac{3.5}{12}\\ &= \frac{3.5}{12}\\ &\approx 0.3\end{aligned}$$

| Weight | Lower | Upper | Mid | Freq. |
|---|---|---|---|---|
| 4.0 - 4.2 | 3.95 | 4.25 | 4.10 | 1 |
| 4.3 - 4.5 | 4.25 | 4.55 | 4.40 | 1 |
| 4.6 - 4.8 | 4.55 | 4.85 | 4.70 | 2 |
| 4.9 - 5.1 | 4.85 | 5.15 | 5.00 | 7 |
| 5.2 - 5.4 | 5.15 | 5.45 | 5.30 | 12 |
| 5.5 - 5.7 | 5.45 | 5.75 | 5.60 | 17 |
| 5.8 - 6.0 | 5.75 | 6.05 | 5.90 | 31 |
| 6.1 - 6.3 | 6.05 | 6.35 | 6.20 | 18 |
| 6.4 - 6.6 | 6.35 | 6.65 | 6.50 | 15 |
| 6.7 - 6.9 | 6.65 | 6.95 | 6.80 | 11 |
| 7.0 - 7.2 | 6.95 | 7.25 | 7.10 | 3 |
| 7.3 - 7.5 | 7.25 | 7.55 | 7.40 | 2 |

| Weight | Freq. | Lower Than | Cum. Freq. |
|---|---|---|---|
| 4.0 - 4.3 | 1 | 4.3 | 1 |
| 4.3 - 4.6 | 1 | 4.6 | 2 |
| 4.6 - 4.9 | 2 | 4.9 | 4 |
| 4.9 - 5.2 | 7 | 5.2 | 11 |
| 5.2 - 5.5 | 12 | 5.5 | 23 |
| 5.5 - 5.8 | 17 | 5.8 | 40 |
| 5.8 - 6.1 | 31 | 6.1 | 71 |
| 6.1 - 6.4 | 18 | 6.4 | 89 |
| 6.4 - 6.7 | 15 | 6.7 | 104 |
| 6.7 - 7.0 | 11 | 7.0 | 115 |
| 7.0 - 7.3 | 3 | 7.3 | 118 |
| 7.3 - 7.6 | 2 | 7.6 | 120 |

(c) Construct a frequency polygon.

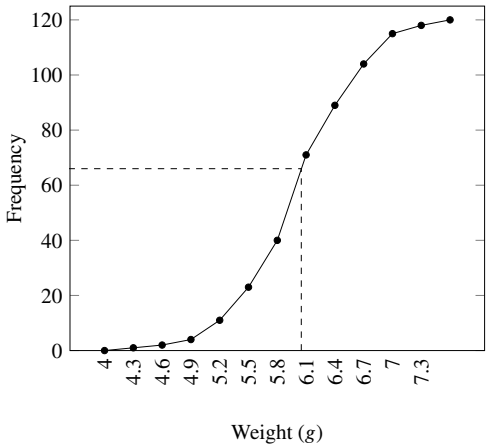
**Sol.**




(d) Construct a cumulative frequency polygon.

**Sol.**

Cumulative Frequency Polygon of Distribution of Weight




(e) Find the percentage of the ears of rice whose length is greater than $6cm$.

**Sol.**

In the diagram above, there are approximately $120-66=54$ ears of rice whose length is greater than $6cm$, which is about $\frac{54}{120}\times 100\% = 45\%$ of the total number of ears of rice.

3. The table below shows the weight distribution of 90 babies (in $kg$):

| Weight | Frequency |
|---|---|
| 1.5 - 2.0 | 2 |
| 2.0 - 2.5 | 4 |
| 2.5 - 3.0 | 13 |
| 3.0 - 3.5 | 32 |
| 3.5 - 4.0 | 28 |
| 4.0 - 4.5 | 10 |
| 4.5 - 5.0 | 1 |

(a) Make a cumulative frequency table.

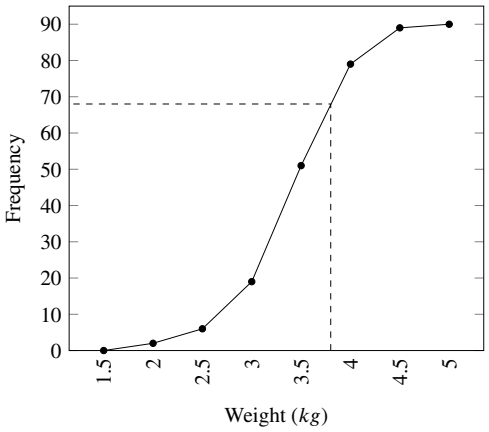
**Sol.**

| Weight | Freq. | Less than | Cum. Freq. |
|---|---|---|---|
| 1.5 - 2.0 | 2 | 2.0 | 2 |
| 2.0 - 2.5 | 4 | 2.5 | 6 |
| 2.5 - 3.0 | 13 | 3.0 | 19 |
| 3.0 - 3.5 | 32 | 3.5 | 51 |
| 3.5 - 4.0 | 28 | 4.0 | 79 |
| 4.0 - 4.5 | 10 | 4.5 | 89 |
| 4.5 - 5.0 | 1 | 5.0 | 90 |

(b) Construct a cumulative frequency polygon.

**Sol.**

Cumulative Frequency Polygon of Distribution of Weight




(c) Find the percentage of babies whose weight is greater than $3.8kg$.

**Sol.**

In the diagram above, there are approximately $90-68=22$ babies whose weight is greater than $3.8kg$, which is about $\frac{22}{90}\times 100\% = 24.44\%$ of the total number of babies.

4. The table below shows the average score distribution of 50 students in a class:

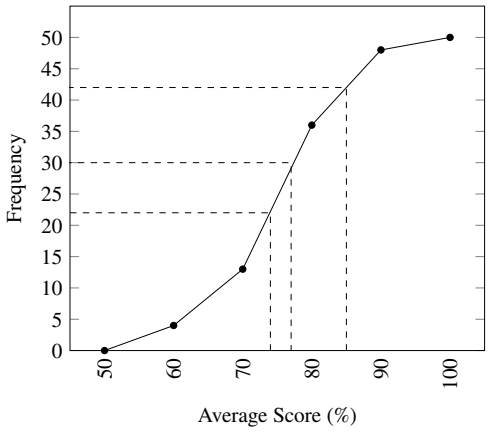
| Average Score | Frequency |
|---|---|
| 50.0 - 59.9 | 4 |
| 60.0 - 69.9 | 9 |
| 70.0 - 79.9 | 23 |
| 80.0 - 89.9 | 12 |
| 90.0 - 99.9 | 2 |

(a) Make a cumulative frequency table and draw a cumulative frequency polygon.

**Sol.**

| Average Score | Freq. | Less than | Cum. Freq. |
|---|---|---|---|
| 50.0 - 59.9 | 4 | 60 | 4 |
| 60.0 - 69.9 | 9 | 70 | 13 |
| 70.0 - 79.9 | 23 | 80 | 36 |
| 80.0 - 89.9 | 12 | 90 | 48 |
| 90.0 - 99.9 | 2 | 100 | 50 |

Cumulative Frequency Polygon of Distribution of Average Score




(b) A student get an average score of 74, find his rank in the class.

**Sol.**

In the diagram above, there are approximately 22 students whose average score is less than 74, which means that the student is ranked $50-22=28$.

(c) Find the average score of the student who is ranked 20.

**Sol.**

In the diagram above, the student who is ranked 20 has an average score of about 77.

(d) Find the percentage of students whose average score is greater than 85.

**Sol.**

In the diagram above, there are approximately $50-42=8$ students whose average score is greater than 85, which is about $\frac{8}{50}\times 100\%=16\%$ of the total number of students.

5. The table below shows the score distribution of 1200 students in UEC accounting exam:

| Score | Number of Students |
|---|---|
| 10 - 19 | 20 |
| 20 - 29 | 60 |
| 30 - 39 | 95 |
| 40 - 49 | 130 |
| 50 - 59 | 340 |
| 60 - 69 | 310 |
| 70 - 79 | 135 |
| 80 - 89 | 80 |
| 90 - 99 | 30 |

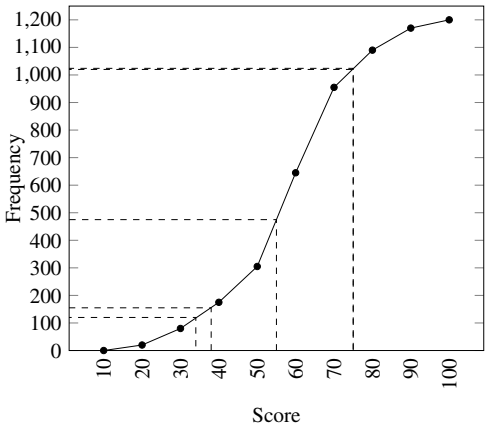
Examinees are categorised into 4 groups based on their score: *Excellent*, *Good*, *Pass*, and *Fail*.

(a) Make a cumulative frequency table and draw a cumulative frequency polygon.

**Sol.**

| Score | Freq. | Less than | Cum. Freq. |
|---|---|---|---|
| 10 - 19 | 20 | 20 | 20 |
| 20 - 29 | 60 | 80 | 80 |
| 30 - 39 | 95 | 175 | 175 |
| 40 - 49 | 130 | 305 | 305 |
| 50 - 59 | 340 | 645 | 645 |
| 60 - 69 | 310 | 955 | 955 |
| 70 - 79 | 135 | 1090 | 1090 |
| 80 - 89 | 80 | 1170 | 1170 |
| 90 - 99 | 30 | 1200 | 1200 |

Cumulative Frequency Polygon of Distribution of Score




(b) If the passing score is 38, find the percentage of students who pass the exam.

**Sol.**

In the diagram above, there are approximately $1200-155=1045$ students whose score is greater or equal to 38, which is about $\frac{1045}{1200}\times 100\%=86.67\%$ of the total number of students.

(c) Assume that the minimum score to be categorised as *Excellent* and *Good* is 75 and 55 respectively, find the percentage of students who are categorised as *Excellent* and *Good* respectively.

**Sol.**

In the diagram above, there are approximately $1200-1024=176$ students whose score is greater or equal to 75, which is about $\frac{176}{1200}\times 100\%=14.67\%$ of the total number of students who are categorised as *Excellent*.

Also, there are approximately $1024 - 475 = 549$ students whose score is greater or equal to 55, which is about $\frac{549}{1200} \times 100\% = 45.75\%$ of the total number of students who are categorised as *Good*.

(d) Find the passing mark if the percentage of students who pass the exam is 90%.

**Sol.**

If the percentage of students who pass the exam is 90%, then the number of students who pass the exam is 90% of 1200 students, which is 1080 students. That means, there are $1200 - 1080 = 120$ students who fail the exam.

In the diagram above, the passing mark is about 34 given that there are 120 students who fail the exam.

(e) Find the minimum mark of a student who is categorised as *Excellent* if the percentage of students who are categorised as *Excellent* is 15%.

**Sol.**

If the percentage of students who are categorised as *Excellent* is 15%, then the number of students who are categorised as *Excellent* is 15% of 1200 students, which is 180 students. That means, there are $1200 - 180 = 1020$ students who are not categorised as *Excellent*.

In the diagram above, the minimum mark of a student who is categorised as *Excellent* is about 75 given that there are 1020 students who are not categorised as *Excellent*.

## 18.3 Central Tendency

Central tendency is a measure of the central position of a distribution, or a single value that attempts to describe a set of data. The most common measures of central tendency are the mean, median, and mode.

### Mean

Mean is also known as arithmetic mean. For $n$ values $x_1, x_2, \ldots, x_n$, the mean is defined as

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 + \cdots + x_n}{n} \\ &= \frac{\sum x_i}{n}\end{aligned}$$

For data whose possible values are $x_1, x_2, \ldots, x_n$, and their respective frequencies are $f_1, f_2, \ldots, f_n$, the mean is defined as

$$\begin{aligned}\bar{x} &= \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{f_1 + f_2 + \cdots + f_n} \\ &= \frac{\sum f_i x_i}{\sum f_i}\end{aligned}$$

For grouped data, we take the mean of each class as the representative value $x_i$ of the class.

#### Weighted Mean

In some scenario, weighted mean is better than the mean to describe the data.

When calculating the arithmetic mean, each value is given equal weight. However, in some cases, each value in a dataset may not be equally important. For example, the importance of the mark of a student for each subject is weighted according to the number of classes of the subject in a week. Hence, when calculating the average mark of the student, each mark must be multiplied by a value that represents the importance of the subject, and that value is called the weight. The weighted mean is defined as

$$\begin{aligned}\bar{x} &= \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} \\ &= \frac{\sum w_i x_i}{\sum w_i}\end{aligned}$$

where $x_i$ are the values and $w_i$ are the weights of $x_i$.

### 18.3.1 Practice 3

1. Find the mean of 34, 50, 24, 32, 53, 30, 62, 27.

   **Sol.**

   $$\begin{aligned}\bar{x} &= \frac{30 + 50 + 24 + 32 + 53 + 30 + 62 + 27}{8} \\ &= \frac{312}{8} \\ &= 39\end{aligned}$$

2. There are three workshop $A$, $B$, and $C$ in a factory. Workshop $A$ has 10 workers, their wages are \$35 per day, workshop $B$ has 30 workers, their wages are \$45 per day, and workshop $C$ has 15 workers, their wages are \$55 per day. Find the mean of the wages of the workers in the factory.

   **Sol.**

   Let the wages of workers be $x_1$, and the amount of workers be $f_1$.

| $x_1$ | $f_1$ | $x_1 f_1$ |
|---|---|---|
| 35 | 10 | 350 |
| 45 | 30 | 1350 |
| 55 | 15 | 825 |
| | $\sum f_i = 55$ | $\sum f_i x_i = 2525$ |

$\therefore$ Average wages of workers in the factory is $\frac{2525}{55} =$ \$45.91.

3. A school appoints students to participate in a Math competition. During the competition, candidates must answer 25 questions within an hour. The table below shows the distribution of frequency of the number of questions that those candidates answer correctly:

| Answered Correctly | Frequency |
|---|---|
| 1 - 5 | 3 |
| 6 - 10 | 12 |
| 11 - 15 | 7 |
| 16 - 20 | 8 |
| 21 - 25 | 5 |

Complete the following table, and find the mean of the number of questions that those candidates answer correctly.

| Ans. Correctly | Freq. $f_i$ | Midpoint $x_i$ | $f_i x_i$ |
|---|---|---|---|
| 1 - 5 | | | |
| 6 - 10 | | | |
| 11 - 15 | | | |
| 16 - 20 | | | |
| 21 - 25 | | | |

**Sol.**

| Ans. Correctly | Freq. $f_i$ | Midpoint $x_i$ | $f_i x_i$ |
|---|---|---|---|
| 1 - 5 | 3 | 3 | 9 |
| 6 - 10 | 12 | 8 | 96 |
| 11 - 15 | 7 | 13 | 91 |
| 16 - 20 | 8 | 18 | 144 |
| 21 - 25 | 5 | 23 | 115 |
| | $\sum f_i = 35$ | $\sum f_i x_i = 455$ | |

$\therefore$ The mean of the number of questions that those candidates answer correctly is $\frac{455}{35} = 13$.

### 18.3.2 Exercise 18.3a

1. Take a sample of 20 from a batch of machine parts, their weight (in $g$) are as follows:

| | | | | | |
|---|---|---|---|---|---|
| 210 | 208 | 200 | 205 | 202 | 218 |
| 206 | 214 | 215 | 207 | 195 | 207 |
| 218 | 192 | 202 | 216 | 185 | 227 |
| 187 | 215 | | | | |

Find the mean weight of these machine parts.

**Sol.**

$$\begin{aligned}\bar{x} &= \frac{210 + 208 + 200 + \cdots + 215}{20} \\ &= \frac{4129}{20} \\ &= 206.45\end{aligned}$$

2. Given that the mean of a dataset 4, -3, 2, $k$, 5, 8 is 10, find the value of $k$.

**Sol.**

$$\begin{aligned}\frac{4 + (-3) + 2 + k + 5 + 8}{6} &= 10 \\ 16 + k &= 60 \\ k &= 44\end{aligned}$$

3. Given that the mean of $x_1$, $x_2$, $x_3$, $x_4$, $x_5$ is 40, and the mean of $y_1$, $y_2$, $y_3$ is 15. Find the mean after combining these two datasets.

**Sol.**

$$\begin{aligned}\frac{x_1 + \cdots + x_5}{5} &= 40 \\ x_1 + \cdots + x_5 &= 200\end{aligned}$$

$$\begin{aligned}\frac{y_1 + y_2 + y_3}{3} &= 15 \\ y_1 + y_2 + y_3 &= 45\end{aligned}$$

$$\begin{aligned}\bar{xy} &= \frac{x_1 + x_2 + \cdots + y_3}{8} \\ &= \frac{245}{8} \\ &= 30.63\end{aligned}$$

4. A school have 2 senior 3 classes: $A$ and $B$. In a Chinese language test, the average mark of 49 students in $A$ class in 72, while the average mark for 45 students

in class $B$ is 68. Find the average mark of all students in these two class combined.

**Sol.**

$$\begin{aligned}\bar{x} &= \frac{72 \times 49 + 68 \times 45}{49 + 45} \\ &= \frac{3528 + 3060}{94} \\ &= \frac{6588}{94} \\ &= 70.09\end{aligned}$$

5. Given that the mean for 8 values are 5. The mean increased by 1.4 after adding two values: $x$ and $3x$. Find the value of $x$.

**Sol.**

$$\begin{aligned}\frac{8 \times 5 + x + 3x}{8 + 2} &= 5 + 1.4 \\ \frac{40 + 4x}{10} &= 6.4 \\ 40 + 4x &= 64 \\ 4x &= 24 \\ x &= 6\end{aligned}$$

6. Throwing 6 coin at the same time and record the number of heads. After throwing 100 times, we get the following frequency distribution table:

| Number of Heads | Frequency |
|---|---|
| 0 | 2 |
| 1 | 10 |
| 2 | 24 |
| 3 | 35 |
| 4 | 22 |
| 5 | 6 |
| 6 | 1 |

Find the mean of the number of heads for each throw.

**Sol.**

Let the number of heads be $x_i$ and the frequency be $f_i$.

| $x_i$ | $f_i$ | $x_i f_i$ |
|---|---|---|
| 0 | 2 | 0 |
| 1 | 10 | 10 |
| 2 | 24 | 48 |
| 3 | 35 | 105 |
| 4 | 22 | 88 |
| 5 | 6 | 30 |
| 6 | 1 | 6 |
| | $\sum f_i = 100$ | $\sum x_i f_i = 287$ |

$\therefore$ The mean of the number of heads for each throw is $\frac{287}{100} = 2.87$.

7. The table below shows the score distribution of 66 students in a Chinese language test:

| Score | Frequency |
|---|---|
| 31 - 40 | 6 |
| 41 - 50 | 12 |
| 51 - 60 | 15 |
| 61 - 70 | 15 |
| 71 - 80 | 8 |
| 81 - 90 | 6 |
| 91 - 100 | 4 |

Find their mark in average.

| Score | Mid $x_1$ | Freq. $f_1$ | $x_1 f_1$ |
|---|---|---|---|
| 31 - 40 | 35.5 | 6 | 213 |
| 41 - 50 | 45.5 | 12 | 546 |
| 51 - 60 | 55.5 | 15 | 832.5 |
| 61 - 70 | 65.5 | 15 | 982.5 |
| 71 - 80 | 75.5 | 8 | 604 |
| 81 - 90 | 85.5 | 6 | 513 |
| 91 - 100 | 95.5 | 4 | 382 |
| | | $\sum f_1 = 66$ | $\sum x_1 f_1 = 4073$ |

$\therefore$ The mark in average is $\frac{4073}{66} = 61.71$.

8. Below are the number of classes and marks for each subject of a junior student:

| Subject | Number of Classes | Average Mark |
|---|---|---|
| Chinese | 7 | 75 |
| Malay | 7 | 73 |
| English | 7 | 65 |
| Mathematics | 7 | 82 |
| Science | 5 | 86 |
| History | 3 | 73 |
| Geography | 3 | 87 |

(a) Find his mark in average.

**Sol.**

$$\begin{aligned}\bar{x} &= \frac{75 + 73 + 65 + 82 + 86 + 73 + 87}{7} \\ &= \frac{541}{7} \\ &= 77.29\end{aligned}$$

(b) Use the number of classes as the weight to find his average mark.

**Sol.**

$$\begin{aligned}\bar{x} &= \frac{75\times 7 + 73\times 7 + \cdots + 87\times 3}{7+7+7+7+5+3+3}\\ &= \frac{525+511+455+574+430+219+261}{39}\\ &= \frac{2975}{39}\\ &= 76.28\end{aligned}$$

9. The weight of 60 junior 2 students in a school are as follows:

| Weight (kg) | Frequency |
|---|---|
| 54 - 56 | 10 |
| 57 - 59 | 20 |
| 60 - 62 | $x$ |
| 63 - 65 | 8 |
| 66 - 68 | 4 |
| 69 - 71 | $y$ |

Given that the mean weight of these students is 60.1 kg, find the value of $x$ and $y$.

**Sol.**

$$\text{Total weight} = 60.1\times 60 = 3606$$

| Wght (kg) | M. $x_1$ | Freq. $f_1$ | $x_1f_1$ |
|---|---|---|---|
| 54 - 56 | 55 | 10 | 550 |
| 57 - 59 | 58 | 20 | 1160 |
| 60 - 62 | 61 | $x$ | $61x$ |
| 63 - 65 | 64 | 8 | 512 |
| 66 - 68 | 67 | 4 | 268 |
| 69 - 71 | 70 | $y$ | $70y$ |
| | | $\sum f_1 = 60$ | $\sum x_1f_1 = 3606$ |

$$\begin{cases} 10+20+x+8+4+y = 60 & (1)\\ 550+1160+61x+512+268+70y = 3606 & (2)\end{cases}$$

$$\begin{aligned}(1) &: 42 + x + y = 60\\ x + y &= 18\\ (2) &: 61x + 70y = 1116\\ (1)\times 61 &: 61x + 61y = 1098\\ (2)-(1) &: 9y = 18\\ y &= 2\\ \text{From (1)} &: x = 16\end{aligned}$$

## Median

The median is the middle value of a sorted dataset. The number of values must be equal for both side of the median.

If the number of values is $n$, when $n$ is odd, the median is the number in $\frac{n+1}{2}$ position.
When $n$ is even, the median is the mean of the number in $\frac{n}{2}$ and $\frac{n}{2}+1$ position.

For grouped data, we can make a cumulative frequency polygon, and the median is the value corresponding to 50% of the percentage of the cumulative frequency.

Let $n$ be the number of values in the dataset, aka $\sum f_1$,

$L_m$ be the lower boundaries of the group of the median,

$C_m$ be the range of the group of the median,

$f_m$ be the frequency of the group of the median,

$F_m$ be the cum. frequency of the group of the median,

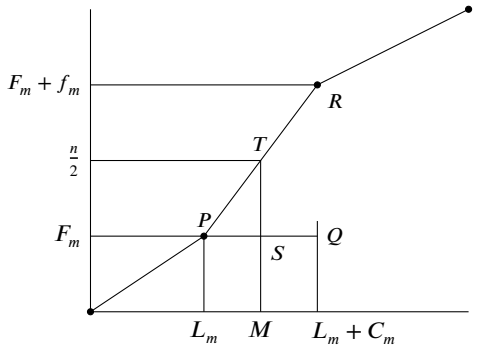



Diagram above shows a part of a cumulative frequency polygon, where $R$ is the point corresponding to the group containing the median, $P$ is the point corresponding to the group before the group containing the median, and $M$ is the median. Since $\Delta PQR \sim \Delta PST$,

$$\therefore \frac{PS}{PQ} = \frac{ST}{QR}$$

$$\text{That is, } \frac{M - L_m}{C_m} = \frac{\frac{n}{2} - F_m}{f_m}$$

We get the following after simplifying the equation:

$$M = L_m + \left(\frac{\frac{n}{2} - F_m}{f_m}\right) C_m$$

### 18.3.3 Practice 4

1. 10 workers in a factory made the same type of product in a day, the number of products made are as follows:

   | | | | | |
   |---|---|---|---|---|
   | 15 | 17 | 14 | 10 | 15 |
   | 19 | 17 | 16 | 14 | 12 |

   Find the median of the number of products made by these 10 workers.

   **Sol.**

   Sort the dataset:

   10 12 14 14 15 15 16 17 17 19

   The median is the mean of the number in $\frac{10}{2} = 5$ and $\frac{10}{2} + 1 = 6$ position, which is $\frac{15+15}{2} = 15$.

2. The table below shows the result of a right eye vision test for 49 students in a class:

   | Vision | Number of Students |
   |---|---|
   | 0.2 | 2 |
   | 0.3 | 3 |
   | 0.4 | 4 |
   | 0.5 | 3 |
   | 0.6 | 4 |
   | 0.8 | 9 |
   | 1.0 | 9 |
   | 1.2 | 10 |
   | 1.5 | 5 |

   Find the median of the right eye vision of these students.

   **Sol.**

   | Vision | Number of Students | Cum. Frequency |
   |---|---|---|
   | 0.2 | 2 | 2 |
   | 0.3 | 3 | 5 |
   | 0.4 | 4 | 9 |
   | 0.5 | 3 | 12 |
   | 0.6 | 4 | 16 |
   | 0.8 | 9 | 25 |
   | 1.0 | 9 | 34 |
   | 1.2 | 10 | 44 |
   | 1.5 | 5 | 49 |

   Since $n = 49$ is odd, the median is the number in the $\frac{49+1}{2} = 25$ position, which is 0.8.

3. The table below shows time distribution of 21 students browsing the Internet:

   | Time (hours) | Number of Students |
   |---|---|
   | 1.1 - 1.3 | 4 |
   | 1.4 - 1.6 | 3 |
   | 1.7 - 1.9 | 5 |
   | 2.0 - 2.2 | 4 |
   | 2.3 - 2.5 | 5 |

   Find the median of the time distribution of these students.

   **Sol.**

   | Time | Freq. | Cum. Freq. |
   |---|---|---|
   | 1.1 - 1.3 | 4 | 4 |
   | 1.4 - 1.6 | 3 | 7 |
   | 1.7 - 1.9 | 5 | 12 |
   | 2.0 - 2.2 | 4 | 16 |
   | 2.3 - 2.5 | 5 | 21 |

   The median is the number in the $\frac{21}{2} = 10.5$ position, which is $1.7 - 1.9$. $C_m = 0.3$, $L_m = 1.65$, and $f_m = 5$, $F_m = 7$.

   $$\therefore Mean = 1.65 + \frac{10.5 - 7}{5} \times 0.3 = 1.86$$

### 18.3.4 Exercise 18.3b

1. During a gymnastic competition, there are four judges scoring the performance of each contestant, and the median of these four scores are taken as the final score of the contestant. Given that the scores given by four judges are 9.5, 9.4, 9.8, and 9.4 respectively, find the final score of the contestant.

   **Sol.**

   Sort the scores:

   9.4 9.4 9.5 9.8

   The median is the mean of the number in $\frac{4}{2} = 2$ and $\frac{4}{2} + 1 = 3$ position, which is $\frac{9.4+9.5}{2} = 9.45$.

2. Following are the weight of 15 boys with same age:

   | | | | | |
   |---|---|---|---|---|
   | 36 | 35 | 33 | 37 | 35 |
   | 42 | 40 | 38 | 38 | 39 |
   | 40 | 41 | 36 | 38 | 37 |

   (a) Find the median of these 15 boys.

   **Sol.**

Sort the data:

33 35 35 36 36 37 37 38 38 38
39 40 41 42 43

The median is the mean of the number in $\frac{15+1}{2} =$ 8 position, which is 38.

(b) Group the data using pattern $33-35, 35-37, \ldots, 41-43$. Then, find the median.

**Sol.**

| Weight ($kg$) | Frequency | Cum. Frequency |
|---|---|---|
| 33 - 35 | 1 | 1 |
| 35 - 37 | 4 | 5 |
| 37 - 39 | 5 | 10 |
| 39 - 41 | 2 | 12 |
| 41 - 43 | 2 | 14 |
| 43 - 45 | 1 | 15 |

The median is the number in the $\frac{15}{2} = 7.5$ position. $C_m = 2$, $L_m = 37$, $f_m = 5$, and $F_m = 5$.

$$\therefore Median = 37 + \frac{7.5-5}{5} \times 2 = 38$$

3. The table below shows the score distribution of a group of pupils in a minor test:

| Score | Number of Pupils |
|---|---|
| 5 | 4 |
| 10 | 2 |
| 15 | 3 |
| 20 | $x$ |
| 25 | 4 |

Assume that the median is 15, find the possibility value of $x$.

**Sol.**

| Score | Freq. | Cum. Freq. |
|---|---|---|
| 5 | 4 | 4 |
| 10 | 2 | 6 |
| 15 | 3 | 9 |
| 20 | $x$ | 9 + x |
| 25 | 4 | 13 + x |

$$\frac{13+x+1}{2} \leq 9$$
$$14 + x \leq 18$$
$$x \leq 4$$
$$\therefore 0 \leq x \leq 4$$

Therefore, the possibility values of $x$ are 0, 1, 2, 3, and 4.

4. The following table shows the income of employees in a company:

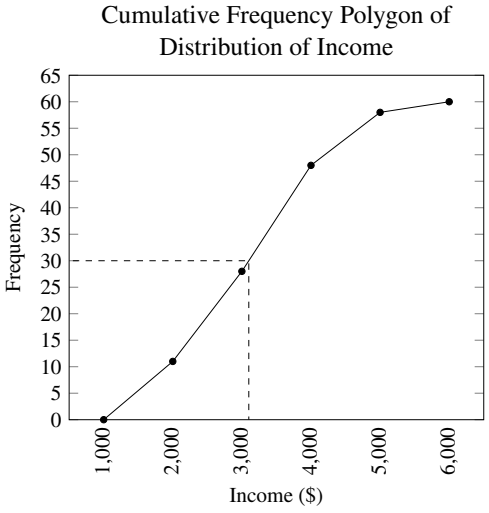
| Income ($) | Number of Employees |
|---|---|
| 1000 - 2000 | 11 |
| 2000 - 3000 | 17 |
| 3000 - 4000 | 20 |
| 4000 - 5000 | 10 |
| 5000 - 6000 | 2 |

(a) Find the median of their income using cumulative frequency polygon.

**Sol.**

| Income ($) | Freq. | Cum. Freq. |
|---|---|---|
| 1000 - 2000 | 11 | 11 |
| 2000 - 3000 | 17 | 28 |
| 3000 - 4000 | 20 | 48 |
| 4000 - 5000 | 10 | 58 |
| 5000 - 6000 | 2 | 60 |

The median is the number in $\frac{60}{2} = 30$ position.




Therefore, the median of their income is \$3100.

(b) Find the median of their income using formula and compare the result with (a).

**Sol.**

The median is the number in the $\frac{60}{2} = 30$ position, which is 3000 − 4000. $C_m = 1000$, $L_m = 3000$, and $f_m = 20$, $F_m = 28$.

$$\therefore Median = 3000 + \frac{30-28}{20} \times 1000 = 3100$$

Therefore, the median of their income is \$3100, which is the same as (a).

5. The table below shows the distribution of height of 20 students:

| Height (cm) | Number of Students |
|---|---|
| 120 - 130 | 3 |
| 130 - 140 | 4 |
| 140 - 150 | $x$ |
| 150 - 160 | 5 |
| 160 - 170 | 6 |

Find:

(a) The value of $x$.

**Sol.**

$$\begin{aligned} x+3+4+5+6 &= 20 \\ x &= 20-18 \\ &= 2 \end{aligned}$$

(b) The median of their height.

**Sol.**

| Height ($cm$) | Freq. | Cum. Freq. |
|---|---|---|
| 120 - 130 | 3 | 3 |
| 130 - 140 | 4 | 7 |
| 140 - 150 | 2 | 9 |
| 150 - 160 | 5 | 14 |
| 160 - 170 | 6 | 20 |

The median is the number in $\frac{20}{2} = 10$ position, which is 150−160. $C_m = 10$, $L_m = 150$, $f_m = 5$, and $F_m = 9$.

$$\therefore Median = 150 + \frac{10-9}{5} \times 10 = 152$$

Therefore, the median of their height is $152cm$.

6. The table below shows the distribution of wages of workers in a factory:

| Wages \$ | Number of Workers |
|---|---|
| 40 - 49 | 4 |
| 50 - 59 | 14 |
| 60 - 69 | 5 |
| 70 - 79 | $x$ |
| 80 - 89 | 2 |

Given that the median is 63.5, find the value of $x$.

**Sol.**

| Wages \$ | Freq. | Cum. Freq. |
|---|---|---|
| 40 - 49 | 4 | 4 |
| 50 - 59 | 14 | 18 |
| 60 - 69 | 5 | 23 |
| 70 - 79 | x | 23 + x |
| 80 - 89 | 2 | 25 + x |

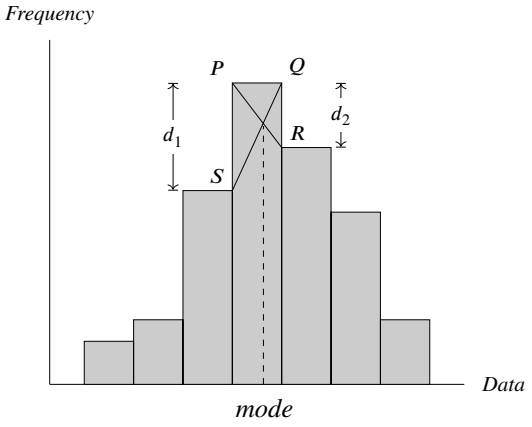
63.5 is in between $60-69$, which is in the $\frac{25+x}{2}$ position. $C_m = 10$, $L_m = 59.5$, $f_m = 5$, $F_m = 18$.

$$\begin{aligned} 59.5 + \frac{\frac{25+x}{2} - 18}{5} \times 10 &= 63.5 \\ \frac{\frac{25+x}{2} - 18}{5} \times 10 &= 4 \\ \frac{\frac{25+x}{2} - 18}{5} &= 0.4 \\ \frac{25+x}{2} - 18 &= 2 \\ \frac{25+x}{2} &= 20 \\ 25 + x &= 40 \\ x &= 15 \end{aligned}$$

## Mode

In a set of data, the mode is the value that occurs most frequently. There can be more than one mode in a set of data. If all the values in a dataset occur with the same frequency, then there is no mode for the data.

For grouped data, the mode is the class that has the highest frequency, and there can be more than one mode. Besides that, the mode can also be estimated using histogram. The method is as follows:




The diagram above shows a histogram of a set of data. The class corresponding to the highest rectangle is the mode of the data, and the mode is the x-value of the intersection point of $PR$ and $QS$.

Unlike median, the formula of mode can be derived from similar triangles. Let:

$L$ be the lower boundaries of the modal class

$C$ be the range of the modal class

$d_1$ be the difference between the lower boundary of the modal class and the lower boundary of the class immediately before the modal class

$d_2$ be the difference between the lower boundary of the modal class and the lower boundary of the class immediately after the modal class

then

$$mode = L + \left(\frac{d_1}{d_1 + d_2}\right) C$$

### 18.3.5 Practice 5

The following table shows the distribution of the score of 36 students in a Mathematics exam:

| Score | Number of Students |
|---|---|
| 20 - 29 | 2 |
| 30 - 39 | 6 |
| 40 - 49 | 10 |
| 50 - 59 | 12 |
| 60 - 69 | 3 |
| 70 - 79 | 2 |
| 80 - 89 | 1 |

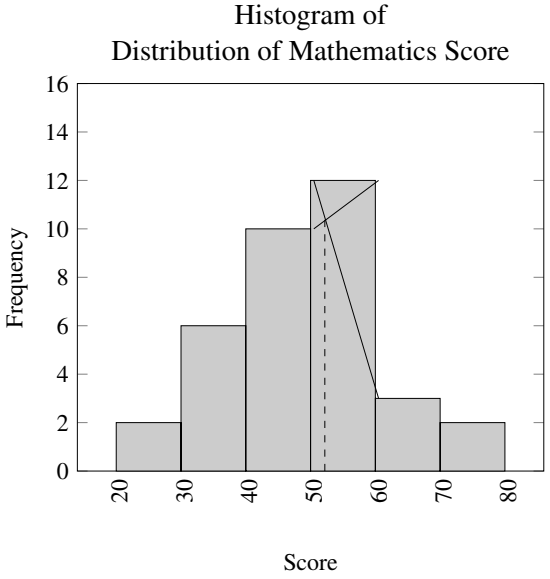
(a) Find the modal class.

**Sol.**

The modal class is $50 - 59$, which has the highest frequency of 12.

(b) Find the mode of score of the students using histogram.

**Sol.**




The mode of score of the students is approximately 51.5.

(c) Find the mode of score of the students using formula.

**Sol.**

$L = 49.5, C = 10, d_1 = 12 - 10 = 2, d_2 = 12 - 3 = 9.$

$$\therefore Mode = 49.5 + \left(\frac{2}{2+9}\right) 10 = 51.32$$

### Comparing mean, median and mode

Generally, the mean, median and mode of a set of data are all dirrent, and they are used to describe the data in different ways.

### 18.3.6 Exercise 18.3c

1. Find the mode of the following data:

   (a) 3 4 3 2 4 5 5 5 4 4

   **Sol.**

   The mode is 4, which has the highest frequency of 4.

   (b) 7 6 8 8 5 6 6 9 8 5

   **Sol.**

   The mode is 6 and 8, which has the highest frequency of 3.

   (c) 1.0 1.1 1.0 0.9 0.8 1.2 1.0 0.9 1.1 1.0

   **Sol.**

   The mode is 1.0, which has the highest frequency of 4.

2. In the sport competition of a high school, the scores of 17 athletes participating in men's high jump are as follows:

| Scores ($m$) | Number of Athletes |
|---|---|
| 1.50 | 2 |
| 1.60 | 3 |
| 1.65 | 2 |
| 1.70 | 3 |
| 1.75 | 4 |
| 1.80 | 1 |
| 1.85 | 1 |
| 1.90 | 1 |

Find the mean, median and mode of their scores.

**Sol.**

$$\begin{aligned} Mean &= \frac{1.50\times 2 + 1.60\times 3 + \cdots + 1.90\times 1}{17} \\ &= \frac{3 + 4.8 + 3.3 + 5.1 + 7 + 1.8 + 1.85 + 1.9}{17} \\ &= \frac{28.75}{17} \\ &= 1.69m \end{aligned}$$

| Scores ($m$) | No. of Athletes | Cum. Frequency |
|---|---|---|
| 1.50 | 2 | 2 |
| 1.60 | 3 | 5 |
| 1.65 | 2 | 7 |
| 1.70 | 3 | 10 |
| 1.75 | 4 | 14 |
| 1.80 | 1 | 15 |
| 1.85 | 1 | 16 |
| 1.90 | 1 | 17 |

The median is the number at $\frac{17+1}{2} = 9$th position, which is $1.70m$.

The mode is $1.75m$, which has the highest frequency of 4.

3. In a Mathematics competition, the scores and the number of students who obtained the scores are as follows:

| Scores (%) | Number of Students |
|---|---|
| 10 - 19 | 20 |
| 20 - 29 | 60 |
| 30 - 39 | 80 |
| 40 - 49 | 40 |
| 50 - 59 | 10 |

Find the modal class and the mode.

**Sol.**

The modal clsas is $30-39$, which has the highest frequency of 80.

$L = 29.5, C = 10, d_1 = 80-60 = 20, d_2 = 80-40 = 40.$

$$\therefore Mode = 29.5 + \left(\frac{20}{20+40}\right)10 = 32.83$$

4. Given that the mean of a dataset 3, 5, 8, 6, 8, 10, 5, 3, $x$, $y$ is 6,

   (a) Prove that $x + y = 12$

   **Proof.**

$$\begin{aligned} \frac{3+5+8+6+8+10+5+3+x+y}{10} &= 6 \\ 48 + x + y &= 60 \\ x + y &= 12 \end{aligned}$$

   (b) With that, if the following conditions are satisfied, find the mode of the dataset.

   i. $x = y$

   **Sol.**

$$\begin{aligned} x &= y \\ x + y &= 12 \\ 2x &= 12 \\ x &= 6 \\ y &= 6 \end{aligned}$$

   $\therefore$ The dataset becomes $3, 5, 8, 6, 8, 10, 5, 3, 6, 6$.
   $\therefore$ The mode is 6, which has the highest frequency of 3.

   ii. $x < y$

   **Sol.**

$$\begin{aligned} x &< y \\ x + y &= 12 \\ 2x &< 12 \\ x &< 6 \\ y &> 6 \end{aligned}$$

   When $x = 1$, $y = 11$, the dataset becomes $3, 5, 8, 6, 8, 10, 5, 3, 1, 11$.
   $\therefore$ The mode are 3, 5, 8, which has the highest frequency of 2.
   When $x = 2$, $y = 10$, the dataset becomes $3, 5, 8, 6, 8, 10, 5, 3, 2, 10$.

$\therefore$ The mode are 3, 5, 8, 10, which has the highest frequency of 2.

When $x = 3$, $y = 9$, the dataset becomes $3, 5, 8, 6, 8, 10, 5, 3, 3, 9$.

$\therefore$ The mode is 3, which has the highest frequency of 3.

When $x = 4$, $y = 8$, the dataset becomes $3, 5, 8, 6, 8, 10, 5, 3, 4, 8$.

$\therefore$ The mode is 8, which has the highest frequency of 3.

When $x = 5$, $y = 7$, the dataset becomes $3, 5, 8, 6, 8, 10, 5, 3, 5, 7$.

$\therefore$ The mode is 5, which has the highest frequency of 3.

5. The mean of a set of data 13, 5, 5, $n$, 5, 10, 10, 11, 9, $n^2$ is 7.4,

   (a) Find the possible values of $n$.

   **Sol.**

$$\frac{13 + 5 + 5 + n + \cdots + n^2}{10} = 7.4$$
$$68 + n + n^2 = 74$$
$$n^2 + n - 6 = 0$$
$$(n + 3)(n - 2) = 0$$
$$n = -3 \text{ or } n = 2$$

   (b) With that, if the following conditions are satisfied, find the median of the dataset.

   i. $n > 0$

   **Sol.**

   If $n > 0$, $n = 2$, the dataset becomes 13, 5, 5, 2, 5, 10, 10, 11, 9, 4. Rearranging the dataset, we get 2, 4, 5, 5, 5, 9, 10, 10, 11, 13. There are a total of 10 elements in the dataset, so the median is the average of the $\frac{10}{2} = 5$th and $\frac{10}{2} + 1 = 6$th elements, which is $\frac{5+9}{2} = 7$.

   ii. $n < 0$

   **Sol.**

   If $n < 0$, $n = -3$, the dataset becomes 13, 5, 5, -3, 5, 10, 10, 11, 9, 9. Rearranging the dataset, we get -3, 5, 5, 5, 9, 9, 10, 10, 11, 13. There are a total of 10 elements in the dataset, so the median is the average of the $\frac{10}{2} = 5$th and $\frac{10}{2} + 1 = 6$th elements, which is $\frac{9+9}{2} = 9$.

6. The following table shows the distribution of scores of a group of students in a competition:

| Scores | Number of Students |
|---|---|
| 0 | 3 |
| 1 | x |
| 2 | 4 |
| 3 | 6 |
| 4 | 2 |

   (a) Assume that the mode is 1, find the minimum value of $x$.

   **Sol.**

   Given that the **only** mode is 1, $x > 6$. Therefore, the minimum value of $x$ is 7.

   (b) Assume that the median is 2, find the maximum value of $x$.

   **Sol.**

   Construct a cumulative frequency table:

| Scores | Number of Students | Cum. Freq. |
|---|---|---|
| 0 | 3 | 3 |
| 1 | $x$ | $3 + x$ |
| 2 | 4 | $7 + x$ |
| 3 | 6 | $13 + x$ |
| 4 | 2 | $15 + x$ |

$$\frac{15 + x}{2} > 3 + x$$
$$15 + x > 6 + 2x$$
$$x < 9$$
$$x = 8$$

   (c) Assume that the mean is 1.95, find the value of $x$.

   **Sol.**

| $x_i$ | $f_i$ | $f_i x_i$ |
|---|---|---|
| 0 | 3 | 0 |
| 1 | $x$ | $x$ |
| 2 | 4 | 8 |
| 3 | 6 | 18 |
| 4 | 2 | 8 |
| | $\sum f_i = 15 + x$ | $\sum = f_i x_i 34 + x$ |

$$\frac{34 + x}{15 + x} = 1.95$$
$$34 + x = 29.25 + 1.95x$$
$$0.95x = 4.75$$
$$x = 5$$

7. Given thet the mode, median and mean of 5 positive integers are 9, 8, and 7.6 respectively, find these 5 numbers.

**Sol.**

Let the 5 numbers be $a, b, c, d, e$, arranged in ascending order.

Since the median is 8, and $c$ is the middle of these 5 numbers, $c = 8$.

Since the mode is 9, there must be more than 1 number that is 9. Since $9 > 8$, $d = e = 9$.

Since the mean is 7.6,

$$\frac{a+b+c+d+e}{5} = 7.6$$
$$\frac{a+b+8+9+9}{5} = 7.6$$
$$a+b+26 = 38$$
$$a+b = 12$$

Since the numbers are arranged in ascending order, and there is only one mode, $a < b < c$, that is, $a < b < 8$.

$$a+b = 12$$
$$a = 12-b$$

$$a < b < 8$$
$$12-b < b < 8$$

$$12 < 2b$$
$$b < 8$$

$$b > 6$$
$$b < 8$$
$$6 < b < 8$$

$$\therefore b = 7$$
$$a = 12-7 = 5$$

Therefore, the 5 numbers are 5, 7, 8, 9, 9.

8. The following table shows the amount of sales of a brand of shoes in a month:

| Shoes Number | Amount of Sales |
|---|---|
| 5 | 4 |
| 6 | 10 |
| 7 | 11 |
| 8 | 18 |
| 9 | 2 |

(a) Find the mean, median, and mode.

**Sol.**

| $x_i$ | $f_i$ | $f_i x_i$ |
|---|---|---|
| 5 | 4 | 20 |
| 6 | 10 | 60 |
| 7 | 11 | 77 |
| 8 | 18 | 144 |
| 9 | 2 | 18 |
| | $\sum f_i = 45$ | $\sum f_i x_i = 319$ |

$$\text{Mean} = \frac{319}{45} = 7.09$$

| Shoes Number | Sales | Cum. Freq. |
|---|---|---|
| 5 | 4 | 4 |
| 6 | 10 | 14 |
| 7 | 11 | 25 |
| 8 | 18 | 43 |
| 9 | 2 | 45 |

Since $n = 45$, the median is the number at the $23^{rd}$ position.

$$\text{Median} = 7$$

Since the number with the highest frequency of 18 is 8,

$$\text{Mode} = 8$$

$\therefore$ mean = 7.09, median = 7, mode = 8.

(b) Which of the following central tendency represents the data best? Why?

**Sol.**

The **mode** is the best central tendency to represent the data because it is the number with the highest frequency, which is 18, and it is the shoes number that most of the shoes are sold.

9. In between 54 examinees in an exam, 15 of them come from cities, 39 of them come from suburbs. Below are the frequency distribution table of their scores:

| Scores | City | Suburb |
|---|---|---|
| 12 - 23 | 0 | 1 |
| 23 - 34 | 0 | 0 |
| 34 - 45 | 0 | 5 |
| 45 - 56 | 1 | 6 |
| 56 - 67 | 3 | 5 |
| 67 - 78 | 4 | 13 |
| 78 - 89 | 6 | 4 |
| 89 - 100 | 1 | 5 |

(a) Find the mean, median, and mode of the scores of the examinees from cities and suburbs respectively.

**Sol.**

For the scores of the examinees from cities:

| Score | Mid. $x_i$ | Freq. $f_i$ | $f_i x_i$ |
|---|---|---|---|
| 12 - 23 | 17.5 | 0 | 0 |
| 23 - 34 | 28.5 | 0 | 0 |
| 34 - 45 | 39.5 | 0 | 0 |
| 45 - 56 | 50.5 | 1 | 50.5 |
| 56 - 67 | 61.5 | 3 | 184.5 |
| 67 - 78 | 72.5 | 4 | 290 |
| 78 - 89 | 83.5 | 6 | 501 |
| 89 - 100 | 94.5 | 1 | 94.5 |
| | | $\sum f_i = 15$ | 1120.5 |

$$\text{Mean} = \frac{1120.5}{15} = 74.7$$

| Score | Feq. | Lower Than | Cum. Freq. |
|---|---|---|---|
| 12 - 23 | 0 | 23 | 0 |
| 23 - 34 | 0 | 34 | 0 |
| 34 - 45 | 0 | 45 | 0 |
| 45 - 56 | 1 | 56 | 1 |
| 56 - 67 | 3 | 67 | 4 |
| 67 - 78 | 4 | 78 | 8 |
| 78 - 89 | 6 | 89 | 14 |
| 89 - 100 | 1 | 100 | 15 |

$n = 15$, $\frac{n}{2} = 7.5$, the group of scores that contains the median is $67 - 78$, $C_m = 11$, $L_m = 67$, $f_m = 3$, $F_m = 4$,

$$\text{Median} = 67 + \frac{7.5 - 4}{4} \times 11 = 76.63$$

Since the modal class is $78 - 89$ with the highest frequency of 6, $C = 11$, $d_1 = 6 - 4 = 2$, $d_2 = 6 - 1 = 5$, $L_m = 78$,

$$\text{Mode} = 78 + \frac{2}{2+5} \times 11 = 81.14$$

$\therefore$ mean = 74.7, median = 76.63, mode = 81.14.

For the scores of the examinees from suburbs:

| Score | Mid. $x_i$ | Freq. $f_i$ | $f_i x_i$ |
|---|---|---|---|
| 12 - 23 | 17.5 | 1 | 17.5 |
| 23 - 34 | 28.5 | 0 | 0 |
| 34 - 45 | 39.5 | 5 | 197.5 |
| 45 - 56 | 50.5 | 6 | 303 |
| 56 - 67 | 61.5 | 5 | 307.5 |
| 67 - 78 | 72.5 | 13 | 942.5 |
| 78 - 89 | 83.5 | 4 | 334 |
| 89 - 100 | 94.5 | 5 | 472.5 |
| | | $\sum f_i = 39$ | 2574.5 |

$$\text{Mean} = \frac{2574.5}{39} = 66.01$$

| Score | Feq. | Lower Than | Cum. Freq. |
|---|---|---|---|
| 12 - 23 | 1 | 23 | 1 |
| 23 - 34 | 0 | 34 | 1 |
| 34 - 45 | 5 | 45 | 6 |
| 45 - 56 | 6 | 56 | 12 |
| 56 - 67 | 5 | 67 | 17 |
| 67 - 78 | 13 | 78 | 30 |
| 78 - 89 | 4 | 89 | 34 |
| 89 - 100 | 5 | 100 | 39 |

$n = 39$, $\frac{n}{2} = 19.5$, the group of scores that contains the median is $67 - 78$, $C_m = 11$, $L_m = 67$, $f_m = 13$, $F_m = 17$,

$$\text{Median} = 67 + \frac{19.5 - 17}{13} \times 11 = 69.12$$

Since the modal class is $67 - 78$ with the highest frequency of 13, $C = 11$, $d_1 = 13 - 5 = 8$, $d_2 = 13 - 4 = 9$, $L = 67$,

$$\text{Mode} = 67 + \frac{8}{8+9} \times 11 = 72.18$$

$\therefore$ mean = 66.01, median = 69.12, mode = 72.18.

Therefore, the mean, median and mode of the scores are:

| | City | Suburbs |
|---|---|---|
| Mean | 74.7 | 66.01 |
| Median | 76.63 | 69.12 |
| Mode | 81.14 | 72.18 |

(b) Find the mean, median and mode of the scores of all the examinees.

**Sol.**

| Score | Mid. $x_i$ | Freq. $f_i$ | $f_i x_i$ |
|---|---|---|---|
| 12 - 23 | 17.5 | 1 | 17.5 |
| 23 - 34 | 28.5 | 0 | 0 |
| 34 - 45 | 39.5 | 5 | 197.5 |
| 45 - 56 | 50.5 | 7 | 353.5 |
| 56 - 67 | 61.5 | 8 | 492 |
| 67 - 78 | 72.5 | 17 | 1232.5 |
| 78 - 89 | 83.5 | 10 | 835 |
| 89 - 100 | 94.5 | 6 | 567 |
| | | $\sum f_i = 54$ | 3695 |

$$\text{Mean} = \frac{3695}{54} = 68.43$$

| Score | Feq. | Lower Than | Cum. Freq. |
|---|---|---|---|
| 12 - 23 | 1 | 23 | 1 |
| 23 - 34 | 0 | 34 | 1 |
| 34 - 45 | 5 | 45 | 6 |
| 45 - 56 | 7 | 56 | 13 |
| 56 - 67 | 8 | 67 | 21 |
| 67 - 78 | 17 | 78 | 38 |
| 78 - 89 | 10 | 89 | 48 |
| 89 - 100 | 6 | 100 | 54 |

$n = 54$, $\frac{n}{2} = 27$, the group of scores that contains the median is $67 - 78$, $C_m = 11$, $L_m = 67$, $f_m = 17$, $F_m = 21$,

$$\text{Median} = 67 + \frac{27 - 21}{17} \times 11 = 70.88$$

Since the modal class is $67 - 78$ with the highest frequency of 17, $C = 11$, $d_1 = 17 - 8 = 9$, $d_2 = 17 - 10 = 7$, $L = 67$,

$$\text{Mode} = 67 + \frac{9}{9 + 7} \times 11 = 73.19$$

10. The following table shows the distribution of scores of a group of students in a Chinese language test:

| Scores $x$ | Number of Students |
|---|---|
| $40 < x \leq 50$ | 12 |
| $50 < x \leq 60$ | 30 |
| $60 < x \leq 70$ | 35 |
| $70 < x \leq 80$ | 25 |
| $80 < x \leq 90$ | 10 |
| $9 < x \leq 100$ | 3 |

Find:

(a) Mean.

**Sol.**

| Scores $x$ | Mid. $x_i$ | Freq. $f_i$ | $f_i x_i$ |
|---|---|---|---|
| $40 < x \leq 50$ | 45 | 12 | 540 |
| $50 < x \leq 60$ | 55 | 30 | 1650 |
| $60 < x \leq 70$ | 65 | 35 | 2275 |
| $70 < x \leq 80$ | 75 | 25 | 1875 |
| $80 < x \leq 90$ | 85 | 10 | 850 |
| $90 < x \leq 100$ | 95 | 3 | 285 |
| | | $\sum f_i = 115$ | 7475 |

$$\text{Mean} = \frac{7475}{115} = 65$$

(b) Modal class and mode.

**Sol.**

The modal class is $60 < x \leq 70$ with the highest frequency of 35, $C = 10$, $d_1 = 35 - 30 = 5$, $d_2 = 35 - 25 = 10$, $L = 60$,

$$\text{Mode} = 60 + \frac{5}{5 + 10} \times 10 = 63.33$$

(c) Median.

**Sol.**

| Score | Freq. | Lower Than | Cum. Freq. |
|---|---|---|---|
| $40 < x \leq 50$ | 12 | 50 | 12 |
| $50 < x \leq 60$ | 30 | 60 | 42 |
| $60 < x \leq 70$ | 35 | 70 | 77 |
| $70 < x \leq 80$ | 25 | 80 | 102 |
| $80 < x \leq 90$ | 10 | 90 | 112 |
| $90 < x \leq 100$ | 3 | 100 | 115 |

$n = 115$, $\frac{n}{2} = 57.5$, the group of scores that contains the median is $60 < x \leq 70$, $C_m = 10$, $L_m = 60$, $f_m = 35$, $F_m = 42$,

$$\text{Median} = 60 + \frac{57.5 - 42}{35} \times 10 = 64.43$$

## 18.4 Measures of Dispersion

The measures of dispersion can be used to describe the spread of the data.

When we're describing a set of data, if we only use the mean, the information provided by the dataset is not enough. For example, given the mean, median, and mode of the average marks of four students in a Mathematics test are all 70 marks, we can't tell the difference between the four students. Their marks might be similar (e.g. 68, 72, 70, 70) or they might be very different (e.g. 100, 40, 70, 70). The latter case is obviously more spread out than the former case.

The most common measures of dispersion are range, interquartile range, quartile deviation, standard deviation, mean deviation, variance, and standard deviation.

## Range

The range of a set of data is the difference between the largest and the smallest value in the dataset.

For grouped data, the range is the difference between the upper limit of the highest class and the lower limit of the lowest class.

## Quartile, Interquartile Range, and Quartile Deviation

Quartiles are three value $Q_1$, $Q_2$, and $Q_3$ that divide a dataset into four equal parts. $Q_2$ is the median of the dataset. $Q_1$ and $Q_3$ are the medians of the two halves of the dataset, called the lower quartile and the upper quartile respectively.

Assume that the number of data in a sorted dataset is $n$. If $n$ is odd, then

When $n$ is even, split the dataset into two halves, with $n/2$ data in each half.

When $n$ is odd, split the data into two halves after removing the median, with $(n-1)/2$ data in each half.

The median of the lower half is $Q_1$ and the median of the upper half is $Q_3$.

For grouped data, we can make a cumulative frequency polygon. In the percentage of the polygon,

25% of the data is below $Q_1$.

50% of the data is below $Q_2$.

75% of the data is below $Q_3$.

Using the same method of deriving the formula for median, we can derive the formula for upper and lower quartiles. Let

$n$ be the number of data in the dataset, aka $\sum f_i$

$L_k$ be the lower boundaries of the class of $Q_k$

$C_k$ be the class range of the class of $Q_k$

$f_k$ be the frequency of the class of $Q_k$

$F_k$ be the cumulative frequency of the class of $Q_k$

then

$$Q_1 = L_1 + \left(\frac{\frac{n}{4} - F_1}{f_1}\right) C_1$$

$$Q_2 = L_2 + \left(\frac{\frac{3n}{4} - F_2}{f_2}\right) C_2$$

The difference between the upper and lower quartiles is called the interquartile range. That is,

$$\textit{Interquartile range} = Q_3 - Q_1$$

The quartile deviation is the interquartile range divided by 2, written as $Q.D.$, that is,

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

Since the interquartile range and the quartile deviation are not affected by the outliers, they are more robust than the range, and are more suitable for representing the spread of the data.

### 18.4.1 Practice 6

1. Find the range, quartiles and interquartile range of the following data:

   (a) 4 8 7 3 3 9 6 5 1 1 2

   **Sol.**

   Sorting the data, we get

   $$\begin{array}{cccccccccccc} 1 & 1 & 2 & 3 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ & & \uparrow & & & \uparrow & & & \uparrow & & \\ & & Q_1 = 2 & & & Q_2 = 4 & & & Q_3 = 7 & & \end{array}$$

   The range is $9 - 1 = 8$.

   The interquartile range is $Q.D. = 7 - 2 = 5$.

   (b) 7 6 8 8 5 6 1 9 8

   **Sol.**

   Sorting the data, we get

   $$\begin{array}{cccccccccc} & & & & Q_2 = 7 & & & & & \\ & & & & \downarrow & & & & & \\ 1 & 5 & 6 & 6 & 7 & 8 & 8 & 8 & 9 & \\ & & \uparrow & & & & \uparrow & & & \\ & & Q_1 = \frac{5+6}{2} = 5.5 & & & & Q_3 = \frac{8+8}{2} = 8 & & & \end{array}$$

   The range is $9 - 1 = 8$.

   The interquartile range is $8 - 5.5 = 2.5$.

   (c) 1.0 1.1 1.5 0.7 0.8 1.2 1.4 0.9 1.6 1.3

   **Sol.**

Sorting the data, we get

$$Q_2 = \frac{1.1+1.2}{2} = 1.15$$

$$\begin{array}{cccccccccc} 0.7 & 0.8 & 0.9 & 1.0 & 1.1 & 1.2 & 1.3 & 1.4 & 1.5 & 1.6 \\ & & \uparrow & & & & & \uparrow & & \\ & & Q_1 = 0.9 & & & & & Q_3 = 1.4 & & \end{array}$$

The range is $1.6 - 0.7 = 0.9$.

The interquartile range is $1.4 - 0.9 = 0.5$.

(d) 3 4 7 2 4 6 5 8

**Sol.**

Sorting the data, we get

$$Q_2 = \frac{4+5}{2} = 4.5$$

$$\begin{array}{cccccccc} 2 & 3 & 4 & 4 & 5 & 6 & 7 & 8 \\ & & \uparrow & & & \uparrow & & \\ & & Q_1 = \frac{3+4}{2} = 3.5 & & & Q_2 = \frac{6+7}{2} = 6.5 & & \end{array}$$

The range is $8 - 2 = 6$.

The interquartile range is $6.5 - 3.5 = 3$.

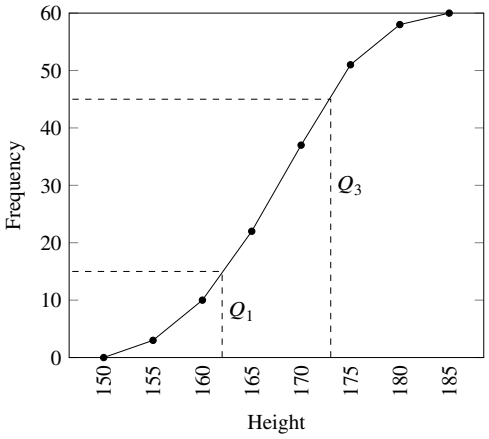
2. The table below shows the cumulative frequency distribution table of the heights of 60 students:

| Height ($cm$) | Cumulative Frequency |
|---|---|
| 150-155 | 3 |
| 155-160 | 10 |
| 160-165 | 22 |
| 165-170 | 37 |
| 170-175 | 51 |
| 175-180 | 58 |
| 180-185 | 60 |

(a) Find the interquartile range of the heights of the students from the cumulative frequency polygon.

**Sol.**

Cumulative Frequency Polygon of Distribution of Height




From the graph, $Q_1$ is approximately $162cm$ and $Q_3$ is approximately $173cm$. Hence,

$$Q.D. = \frac{173 - 162}{2} = 5.5$$

(b) Find the interquartile range of the heights of the students using formula.

**Sol.**

$n = 60$, $\frac{n}{4} = 15$, the class that contains $Q_1$ is $160 - 165$, $C_1 = 5$, $L_1 = 160$, $f_1 = 11$, $F_1 = 10$,

$$Q_1 = 160 + \frac{15 - 10}{12} \times 5 = 162.083$$

$n = 60$, $\frac{3n}{4} = 45$, the class that contains $Q_3$ is $170 - 175$, $C_3 = 5$, $L_3 = 170$, $f_3 = 14$, $F_3 = 37$,

$$Q_3 = 170 + \frac{45 - 37}{14} \times 5 = 172.857$$

Hence,

$$Q.D. = \frac{172.857 - 162.083}{2} = 5.39$$

### 18.4.2 Exercise 18.4a

1. Following are the sales of televisions of a shop in 11 days:

4 9 0 1 3 4 2 5 7 2 3

Find:

(a) The range.

**Sol.**

Rearranging the data,

0 1 2 2 3 3 4 4 5 7 9

Hence, the range is $9 - 0 = 9$.

(b) The quartiles and interquartile range.

**Sol.**

Rearranging the data,

0 1 2 2 3 3 4 4 5 7 9

$Q_1 = 2$ $Q_2 = 3$ $Q_3 = 5$

Hence, the interquartile range is $5 - 2 = 3$.

2. Given a set of data: 1.2, 1.0, 1.1, 1.3, 1.5, 1.7, 1.2, 1.0. Find:

(a) The range.

**Sol.**

Rearranging the data,

1.0 1.0 1.1 1.2 1.2 1.3 1.5 1.7

Hence, the range is $1.7 - 1.0 = 0.7$.

(b) The quartiles and interquartile deviation.

**Sol.**

Rearranging the data,

$Q_2 = \frac{1.2+1.2}{2} = 1.2$

1.0 1.0 1.1 1.2 1.2 1.3 1.5 1.7

$Q_1 = \frac{1.0}{1.1} = 1.05$ $Q_3 = \frac{1.3+1.5}{2} = 1.4$

Hence, the interquartile deviation is $\frac{1.4-1.05}{2} = 0.18$.

3. The distribution of scores of Mathematics test of 100 senior 1 students from a high school are as follows:

| Scores | Number of Students |
|---|---|
| 30 - 40 | 3 |
| 40 - 50 | 4 |
| 50 - 60 | 13 |
| 60 - 70 | 22 |
| 70 - 80 | 30 |
| 80 - 90 | 23 |
| 90 - 100 | 5 |

Find the interquartile deviation of the scores.

**Sol.**

| Scores | Freq. | Lower Than | Cumulative Freq. |
|---|---|---|---|
| 30 - 40 | 3 | 40.5 | 3 |
| 40 - 50 | 4 | 50.5 | 7 |
| 50 - 60 | 13 | 60.5 | 20 |
| 60 - 70 | 22 | 70.5 | 42 |
| 70 - 80 | 30 | 80.5 | 72 |
| 80 - 90 | 23 | 90.5 | 95 |
| 90 - 100 | 5 | 100.5 | 100 |

$n = 100$, $\frac{n}{4} = 25$, The class that contains the lower quartile is $60 - 70$, $C_1 = 10$, $L_1 = 60.5$, $f_1 = 22$, $F_1 = 20$,

$$Q_1 = 60.5 + \frac{25 - 20}{22} \times 10 = 62.773$$

$n = 100$, $\frac{3n}{4} = 75$, The class that contains the upper quartile is $80 - 90$, $C_2 = 10$, $L_2 = 80.5$, $f_2 = 23$, $F_2 = 72$,

$$Q_3 = 80.5 + \frac{75 - 72}{23} \times 10 = 81.804$$

Hence,

$$Q.D. = \frac{81.804 - 62.773}{2} = 9.52$$

**Mean Deviation**

Let the mean of a set of data $x_1, x_2, \ldots, x_n$ be $\bar{x}$, $|x_i - \bar{x}|$ is the difference between the $i$th data and the mean, the mean of these $n$ differences are called the mean deviation, and can be used to calculate the measure of dispersion of the data. That is,

$$Mean\ Deviation = \frac{\sum |x_i - \bar{x}|}{n}$$

If the possible value given data are $x_1, x_2, \ldots, x_n$, their frequencies are $f_1, f_2, \ldots, f_n$, respectively, then the mean deviation can be calculated as follows:

$$Mean\ Deviation = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i}$$

For grouped data, we take the midpoints of the classes as the representative value $x_i$.

### 18.4.3 Practice 7

Complete the following table, and find the mean and mean deviation of the data.

| Lim. | $f_i$ | Mid. $x_i$ | $f_i x_i$ | $\lvert x_i - \bar{x}\rvert$ | $\lvert x_i - \bar{x}\rvert f_i$ |
|---|---|---|---|---|---|
| 50 - 54 | 2 | | | | |
| 55 - 59 | 3 | | | | |
| 60 - 64 | 6 | | | | |
| 65 - 69 | 9 | | | | |

**Sol.**

| Lim. | $f_i$ | Mid. $x_i$ | $f_i x_i$ | $\lvert x_i - \bar{x}\rvert$ | $\lvert x_i - \bar{x}\rvert f_i$ |
|---|---|---|---|---|---|
| 50 - 54 | 2 | 52 | 104 | 10.5 | 21 |
| 55 - 59 | 3 | 57 | 171 | 5.5 | 16.5 |
| 60 - 64 | 6 | 62 | 372 | 0.5 | 3 |
| 65 - 69 | 9 | 67 | 603 | 4.5 | 40.5 |
| | 20 | | 1250 | | 81 |

$$\bar{x} = \frac{1250}{20} = 62.5$$
$$\textit{Mean Deviation} = \frac{81}{20} = 4.05$$

### 18.4.4 Exercise 18.4b

1. Find the mean deviation of the following dataset:

   (a) 7 10 9 12 4 11 3

   **Sol.**

   $$\bar{x} = \frac{7+10+9+12+4+11+3}{7} = 8$$
   $$\textit{Mean Dev.} = \frac{1}{7}(|7-8| + \cdots + |3-8|)$$
   $$= 2.86$$

   (b) 58 65 38 76 43

   **Sol.**

   $$\bar{x} = \frac{58+65+38+76+43}{5} = 56$$
   $$\textit{Mean Dev.} = \frac{1}{5}(|58-56| + \cdots + |43-56|)$$
   $$= 12.4$$

   (c) 45.0 46.5 47.0 48.0 48.7 48.9 49.5 50.4

   **Sol.**

   $$\bar{x} = \frac{45.0+46.5+\cdots+50.4}{8} = 48$$
   $$\textit{M Dev.} = \frac{1}{8}(|45.0-48.3| + \cdots + |50.4-48.3|)$$
   $$= 1.38$$

2. The table below shows the frequency of the number of questions answered correctly by 26 students in a Mathematics minor test:

| Num. of Corr. Ans. Ques. | Num. of Stud. |
|---|---|
| 1 | 0 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 6 |
| 6 | 8 |
| 7 | 6 |
| 8 | 1 |
| 9 | 1 |
| 10 | 1 |

   Find the mean deviation fo the number of questions answered correctly.

   **Sol.**

| $x_i$ | $f_i$ | $f_i x_i$ | $\lvert x_i - \bar{x}\rvert$ | $\lvert x_i - \bar{x}\rvert f_i$ |
|---|---|---|---|---|
| 1 | 0 | 0 | 5 | 0 |
| 2 | 1 | 2 | 4 | 4 |
| 3 | 1 | 3 | 3 | 3 |
| 4 | 1 | 4 | 2 | 2 |
| 5 | 6 | 30 | 1 | 6 |
| 6 | 8 | 48 | 0 | 0 |
| 7 | 6 | 42 | 1 | 6 |
| 8 | 1 | 8 | 2 | 2 |
| 9 | 1 | 9 | 3 | 3 |
| 10 | 1 | 10 | 4 | 4 |
| | 26 | | | 30 |

   $$\bar{x} = \frac{156}{26} = 6$$
   $$\textit{Mean Deviation} = \frac{30}{26} = 1.15$$

3. Following are the test scores of 36 students:

   77 60 52 73 60 50 70 60 52
   68 59 50 72 59 48 66 58 46
   60 48 34 61 55 40 62 55 42
   63 55 43 65 56 45 65 57 46

   (a) Group the dataset above according to the pattern $[34-38), [38-42), [42-46), \ldots$, then make a frequency distribution table.

   **Sol.**

| Range | Frequency |
|---|---|
| $34 \le x < 38$ | 1 |
| $38 \le x < 42$ | 1 |
| $42 \le x < 46$ | 3 |
| $46 \le x < 50$ | 4 |
| $50 \le x < 54$ | 4 |
| $54 \le x < 58$ | 5 |
| $58 \le x < 62$ | 8 |
| $62 \le x < 66$ | 4 |
| $66 \le x < 70$ | 2 |
| $70 \le x < 74$ | 3 |
| $74 \le x < 78$ | 1 |

(b) Find the mean from the frequency distribution table.

**Sol.**

| Range | Mid $x_i$ | Freq. $f_i$ | $f_i x_i$ |
|---|---|---|---|
| $34 \le x < 38$ | 36 | 1 | 36 |
| $38 \le x < 42$ | 40 | 1 | 40 |
| $42 \le x < 46$ | 44 | 3 | 132 |
| $46 \le x < 50$ | 48 | 4 | 192 |
| $50 \le x < 54$ | 52 | 4 | 208 |
| $54 \le x < 58$ | 56 | 5 | 280 |
| $58 \le x < 62$ | 60 | 8 | 480 |
| $62 \le x < 66$ | 64 | 4 | 256 |
| $66 \le x < 70$ | 68 | 2 | 136 |
| $70 \le x < 74$ | 72 | 3 | 216 |
| $74 \le x < 78$ | 76 | 1 | 76 |
| | | 36 | 2052 |

$$\bar{x} = \frac{2052}{36} = 57$$

(c) Find the mean deviation from the frequency distribution table.

**Sol.**

| $x_i$ | $f_i$ | $\lvert x_i - \bar{x}\rvert$ | $f_i \lvert x_i - \bar{x}\rvert$ |
|---|---|---|---|
| 36 | 1 | 21 | 21 |
| 40 | 1 | 17 | 17 |
| 44 | 3 | 13 | 39 |
| 48 | 4 | 9 | 36 |
| 52 | 4 | 5 | 20 |
| 56 | 5 | 1 | 5 |
| 60 | 8 | 3 | 24 |
| 64 | 4 | 7 | 28 |
| 68 | 2 | 11 | 22 |
| 72 | 3 | 15 | 45 |
| 76 | 1 | 19 | 19 |
| | 36 | 107 | 276 |

$$\textit{Mean Deviation} = \frac{276}{36} = 7.67$$

**Variance and Standard Deviation**

Let the mean of a set of data $x_1, x_2, \ldots, x_n$ be $\bar{x}$, $(x_1 - \bar{x})^2$ be the square of the difference between the $i^{th}$ data and the mean, the square of the mean of these $n$ differences are called the variance, written as $\sigma^2$, that is,

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

The square root of the variance is called the standard deviation, written as $\sigma$, that is,

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

If the possible values of given data are $x_1, x_2, \ldots, x_n$, their frequencies are $f_1, f_2, \ldots, f_n$, respectively, then

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$$

For grouped data, we take the midpoints of the classes as the representative value $x_i$.

The above formula are a bit complicated, so we can simplify the formula:

$$\begin{aligned}
\sigma^2 &= \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} \\
&= \frac{\sum x_i^2 f_i - 2\bar{x} \sum x_i f_i + \sum \bar{x}^2 f_i}{\sum f_i} \\
&= \frac{\sum x_i^2 f_i}{\sum f_i} - 2\bar{x}^2 + \bar{x}^2 \\
&= \frac{\sum x_i^2 f_i}{\sum f_i} - \bar{x}^2
\end{aligned}$$

Hence, when the frequency of value $x_i$ is $f_i$, Then

$$\sigma^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - \bar{x}^2$$

$$\sigma = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i} - \bar{x}^2}$$

When all the frequencies $f_i$ are equal to 1, then

$$\sigma^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

$$\sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

Compared to mean deviation, the variance and standard deviation do not contain absolute value, so it is more convenient to use them. Furthermore, the variance and standard deviation are more sensitive to the difference between the data and the mean, so they are more commonly used in daily life.

### 18.4.5 Practice 8

1. Measuring the height of 10 plant seedlings (in $cm$) in a lab, we get the following data:

   12 6 15 3 12 6 21 15 18 12

   Find the standard deviation of the height of the plant seedlings.

   **Sol.**

$$\begin{aligned}\bar{x} &= \frac{12+6+\cdots+12}{10} = 12\\ \sigma^2 &= \frac{12^2+6^2+\cdots+12^2}{10} - 12^2\\ &= \frac{1728}{10} - 144\\ &= 28.8\\ \sigma &= \sqrt{28.8} = 5.37\end{aligned}$$

2. Complete the following table, then find the standard deviation.

   (a)

| $x_i$ | $f_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|
| 3 | 30 | | |
| 5 | 35 | | |
| 7 | 28 | | |

   **Sol.**

| $x_i$ | $f_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|
| 3 | 30 | 90 | 270 |
| 5 | 35 | 175 | 875 |
| 7 | 28 | 196 | 1372 |
| | $\sum f_i = 93$ | $\sum x_i f_i = 461$ | $\sum x_i^2 f_i = 2517$ |

$$\begin{aligned}\bar{x} &= \frac{461}{93} = 4.96\\ \sigma^2 &= \frac{2517}{93} - 4.96^2\\ &= 27.06 - 24.60\\ &= 2.46\\ \sigma &= \sqrt{2.46} = 1.57\end{aligned}$$

   (b)

| Limit | $f_i$ | $Mid.x_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|---|
| 150 - 154 | 5 | | | |
| 155 - 159 | 8 | | | |
| 160 - 164 | 10 | | | |
| 165 - 169 | 7 | | | |
| 170 - 174 | 6 | | | |
| 175 - 179 | 4 | | | |

   **Sol.**

| Limit | $f_i$ | $Mid.x_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|---|
| 150 - 154 | 5 | 152 | 760 | 115520 |
| 155 - 159 | 8 | 157 | 1256 | 197192 |
| 160 - 164 | 10 | 162 | 1620 | 262440 |
| 165 - 169 | 7 | 167 | 1169 | 195223 |
| 170 - 174 | 6 | 172 | 1032 | 177504 |
| 175 - 179 | 4 | 177 | 708 | 125316 |
| | $\sum f_i = 40$ | | $\sum x_i f_i = 6545$ | $\sum x_i^2 f_i = 1073195$ |

$$\begin{aligned}\bar{x} &= \frac{6545}{40} = 163.625\\ \sigma^2 &= \frac{1073195}{40} - 163.88^2\\ &= 26829.88 - 26773.14\\ &= 56.74\\ \sigma &= \sqrt{55.74} = 7.53\end{aligned}$$

### 18.4.6 Exercise 18.4c

1. Find the variance and standard deviation of the following dataset:

   (a) 3 6 3 8

   **Sol.**

$$\begin{aligned}\bar{x} &= \frac{3+6+3+8}{4} = 5\\ \sigma^2 &= \frac{3^2+6^2+3^2+8^2}{4} - 5^2\\ &= 29.5 - 25\\ &= 4.5\\ \sigma &= \sqrt{4.5} = 2.12\end{aligned}$$

   (b) 3 3 4 5 10

   **Sol.**

$$\begin{aligned}\bar{x} &= \frac{3+3+4+5+10}{5} = 5\\ \sigma^2 &= \frac{3^2+3^2+4^2+5^2+10^2}{5} - 5^2\\ &= 31.8 - 25\\ &= 6.8\\ \sigma &= \sqrt{6.8} = 2.61\end{aligned}$$

(c) 2 9 10 10 12 2 10 9

**Sol.**

$$\bar{x} = \frac{2+9+\cdots+10+9}{8} = 8$$
$$\sigma^2 = \frac{2^2+9^2+\cdots+9^2}{8} - 8^2$$
$$= 76.75 - 64$$
$$= 12.75$$
$$\sigma = \sqrt{12.75} = 3.57$$

2. Find the variance and standard deviation of the data:

(a)

| Values | Frequency |
|---|---|
| 6 | 35 |
| 5 | 36 |
| 4 | 30 |

**Sol.**

| $x_i$ | $f_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|
| 6 | 35 | 210 | 1260 |
| 5 | 36 | 180 | 900 |
| 4 | 30 | 120 | 480 |
| | $\sum f_i = 101$ | $\sum x_i f_i = 510$ | $\sum x_i^2 f_i = 2640$ |

$$\bar{x} = \frac{510}{101} = 5.05$$
$$\sigma^2 = \frac{2640}{101} - 5.05^2$$
$$= 26.14 - 25.5$$
$$= 0.64$$
$$\sigma = \sqrt{0.64} = 0.80$$

(b)

| Values | Frequency |
|---|---|
| 60 | 4 |
| 70 | 6 |
| 80 | 2 |
| 90 | 5 |
| 100 | 1 |

**Sol.**

| $x_i$ | $f_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|
| 60 | 4 | 240 | 14400 |
| 70 | 6 | 420 | 29400 |
| 80 | 2 | 160 | 12800 |
| 90 | 5 | 450 | 40500 |
| 100 | 1 | 100 | 10000 |
| | $\sum f_i = 18$ | $\sum x_i f_i = 1370$ | $\sum x_i^2 f_i = 107100$ |

$$\bar{x} = \frac{1370}{18} = 76.11$$
$$\sigma^2 = \frac{107100}{18} - 76.11^2$$
$$= 5950 - 5792.90$$
$$= 157.10$$
$$\sigma = \sqrt{157.10} = 12.53$$

3. Given two sets of data:

| A | B |
|---|---|
| 9.9 | 10.3 |
| 10.3 | 10 |
| 9.8 | 9.5 |
| 10.1 | 10.4 |
| 10.4 | 10.5 |
| 10 | 9.4 |
| 9.8 | 9.8 |
| 9.7 | 10.1 |

Find the mean and variance of these two sets of data respectively, and state which set of data is more spread out.

**Sol.**

For data $A$,

$$\bar{x} = \frac{9.9+10.3+\cdots+9.7}{8} = 10$$
$$\sigma^2 = \frac{9.9^2+10.3^2+\cdots+9.7^2}{8} - 10^2$$
$$= 100.055 - 100$$
$$= 0.06$$

For data $B$,

$$\bar{x} = \frac{10.3+10+\cdots+10.1}{8} = 10.0$$
$$\sigma^2 = \frac{10.3^2+10^2+\cdots+10.1^2}{8} - 10.05^2$$
$$= 100.145 - 100$$
$$= 0.15$$

Since $\sigma^2$ of data $B$ is larger than that of data $A$, data $B$ is more spread out.

4. Given the Chinese language test scores of two groups of students are as follows:

| Group A | Group B |
|---|---|
| 76 | 82 |
| 90 | 84 |
| 84 | 85 |
| 86 | 89 |
| 81 | 79 |
| 87 | 80 |
| 86 | 91 |
| 82 | 89 |
| 85 | 79 |
| 83 | 74 |

Find the mean and standard deviation of these two sets of data respectively, and state which set of data is more centered.

**Sol.**

For Group $A$,

$$\begin{aligned}\bar{x} &= \frac{76+90+\cdots+83}{10} = 84\\ \sigma^2 &= \frac{76^2+90^2+\cdots+83^2}{10} - 84^2\\ &= 7069.2 - 7056\\ &= 13.2\\ \sigma &= \sqrt{13.2} = 3.63\end{aligned}$$

For Group $B$,

$$\begin{aligned}\bar{x} &= \frac{82+84+\cdots+74}{10} = 83.2\\ \sigma^2 &= \frac{82^2+84^2+\cdots+74^2}{10} - 83.2^2\\ &= 6948.6 - 6922.24\\ &= 26.36\\ \sigma &= \sqrt{26.36} = 5.13\end{aligned}$$

Since $\sigma$ of Group $A$ is smaller than that of Group $B$, Group $A$ is more centered.

5. The table below shows the height distribution of all students of the same grade:

| Height ($cm$) | Frequency |
|---|---|
| 145 - 149 | 10 |
| 150 - 154 | 36 |
| 155 - 159 | 193 |
| 160 - 164 | 205 |
| 165 - 169 | 240 |
| 170 - 174 | 83 |
| 175 - 179 | 33 |

Find the mean and standard deviation of the height of all students of the same grade.

**Sol.**

| Range | $x_i$ | $f_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|---|
| 145 - 149 | 147 | 10 | 1470 | 216090 |
| 150 - 154 | 152 | 36 | 5472 | 831744 |
| 155 - 159 | 157 | 193 | 30301 | 4757257 |
| 160 - 164 | 162 | 205 | 33210 | 5380020 |
| 165 - 169 | 167 | 240 | 40080 | 6693360 |
| 170 - 174 | 172 | 83 | 14276 | 2455472 |
| 175 - 179 | 177 | 33 | 5841 | 1033857 |
| | | $\sum f_i = 800$ | $\sum f_i x_i = 130650$ | $\sum f_i x_i^2 = 21367800$ |

$$\begin{aligned}\bar{x} &= \frac{130650}{800} = 163.31cm\\ \sigma^2 &= \frac{21367800}{800} - 163.31^2\\ &= 26709.75 - 26670.16\\ &= 38.78cm\\ \sigma &= \sqrt{38.78} = 6.23cm\end{aligned}$$

6. Following are teh weight distribution of 100 students in a school:

| Weight ($kg$) | Number of Students |
|---|---|
| 45 - 47 | 3 |
| 48 - 50 | 16 |
| 51 - 53 | 20 |
| 54 - 56 | 32 |
| 57 - 59 | 15 |
| 60 - 62 | 10 |
| 63 - 65 | 4 |

Find the variance and standard deviation.

**Sol.**

| Range | $x_i$ | $f_i$ | $x_i f_i$ | $x_i^2 f_i$ |
|---|---|---|---|---|
| 45 - 47 | 46 | 3 | 138 | 6348 |
| 48 - 50 | 49 | 16 | 784 | 38416 |
| 51 - 53 | 52 | 20 | 1040 | 54080 |
| 54 - 56 | 55 | 32 | 1760 | 96800 |
| 57 - 59 | 58 | 15 | 870 | 50460 |
| 60 - 62 | 61 | 10 | 610 | 37210 |
| 63 - 65 | 64 | 4 | 256 | 16384 |
| | | $\sum f_i = 100$ | $\sum f_i x_i = 5458$ | $\sum f_i x_i^2 = 299698$ |

$$\begin{aligned}\bar{x} &= \frac{5458}{100} = 54.58kg\\ \sigma^2 &= \frac{299698}{100} - 54.58^2\\ &= 2996.98 - 2978.98\\ &= 18.00kg\\ \sigma &= \sqrt{18.00} = 4.24kg\end{aligned}$$

7. Given the sum of 10 values is 400, and the sum of their square is 16400. Find the mean and variance of these 10 values.

**Sol.**

$$\begin{aligned}\bar{x} &= \frac{400}{10} = 40\\ \sigma^2 &= \frac{16400}{10} - 40^2\\ &= 1640 - 1600\\ &= 40\end{aligned}$$

8. Given 30 values $x_1, x_2, \cdots, x_{30}$, the mean of these values is 5, and the standard deviation is 2. Find $\sum_{i=1}^{30} x_i$ and $\sum_{i=1}^{30} x_i^2$.

**Sol.**

$$\begin{aligned}\bar{x} &= 5\\ \frac{\sum_{i=1}^{30} x_i}{30} &= 5\\ \sum_{i=1}^{30} x_i &= 150\end{aligned}$$

$$\begin{aligned}\sigma &= 2\\ \sigma^2 &= 4\\ \frac{\sum_{i=1}^{30} x_i^2}{30} - 5^2 &= 4\\ \frac{\sum_{i=1}^{30} x_i^2}{30} &= 29\\ \sum_{i=1}^{30} x_i^2 &= 870\end{aligned}$$

9. The mean of 5 values is 10, and it remains the same after adding $p$ to dataset.

   (a) Find the value of $p$.

   **Sol.**

   Let the dataset be $x$.

$$\begin{aligned}\bar{x} &= 10\\ \frac{\sum_{i=1}^{5} x_i}{5} &= 10\\ \sum_{i=1}^{5} x_i &= 50\\ \frac{\sum_{i=1}^{5} x_i + p}{6} &= 10\\ 50 + p &= 60\\ p &= 10\end{aligned}$$

   (b) If the sum of square of 5 original values is 558, find the variance of the 6 values after adding $p$.

   **Sol.**

$$\begin{aligned}\sigma^2 &= \frac{\sum_{i=1}^{5} x_i^2 + p^2}{6} - 10^2\\ &= \frac{558 + 100}{6} - 100\\ &= 109.67 - 100\\ &= 9.67\end{aligned}$$

10. Given that the mean of 3, 6, 7, 8, 9, 12, 14, 15, $x$, $y$ is 13, standard deviation is $\sqrt{102}$, find the value of $x$ and $y$.

**Sol.**

$$\begin{aligned}\bar{x} = \frac{3 + 6 + \cdots + 15 + x + y}{10} &= 13\\ \frac{74 + x + y}{10} &= 13\\ 74 + x + y &= 130\\ x + y &= 56\end{aligned}$$

$$\begin{aligned}\sigma &= \sqrt{102}\\ \sigma^2 &= 102\\ \frac{3^2 + 6^2 + \cdots + 15^2 + x^2 + y^2}{10} - 13^2 &= 102\\ \frac{804 + x^2 + y^2}{10} - 169 &= 102\\ \frac{804 + x^2 + y^2}{10} &= 271\\ 804 + x^2 + y^2 &= 2710\\ x^2 + y^2 &= 1906\end{aligned}$$

$$\begin{cases} x + y = 56 & (1) \\ x^2 + y^2 = 1906 & (2) \end{cases}$$

$$(1) \Rightarrow y = 56 - x \quad (1)$$

$$\text{Sub (1) into (2)} \Rightarrow x^2 + (56 - x)^2 = 1906$$

$$x^2 + x^2 - 112x + 3136 = 1906$$

$$2x^2 - 112x + 1230 = 0$$

$$x^2 - 56x + 615 = 0$$

$$(x - 15)(x - 41) = 0$$

$$x = 15 \text{ or } x = 41$$

When $x = 15$, $y = 56 - 15 = 41$

When $x = 41$, $y = 56 - 41 = 15$

$$\therefore \begin{cases} x = 15 \\ y = 41 \end{cases} \text{ or } \begin{cases} x = 41 \\ y = 15 \end{cases}$$

## 18.5 Coefficient of Variation

Generally speaking, when we want to compare the variability of two or more sets of data, only comparing the standard deviation of each group is not enough. If the properties or the units of the data are different, the standard deviation of each group must not be comparable. For example, if we want to know whether the deviation of the height of students in a class is larger than that of the weight of students in the same class, we need a relative metric as the standard of comparison, and the coefficient of variation is such a metric. For a non-negative set of value, the definition of coefficient of variation is as follows:

$$v = \frac{\sigma}{\bar{x}} \times 100\%$$

From the definition, we can see that the coefficient of variation the standard deviation when the mean is 1. Thus, when the coefficient of variation is large, it means that the variability of the data is large, and vice versa.

### 18.5.1 Practice 9

In a minor test, the full mark of Chinese language test for senior 2 students is 100, its average mark is 70, and the standard deviation is 10, while the full mark of Mathematics test is 70, its average mark is 40, and the standard deviation is 8. Compare the variability of the two tests.

**Sol.**

$$v_{\text{Chinese}} = \frac{10}{70} \times 100\% = 14.29\%$$

$$v_{\text{Math}} = \frac{8}{40} \times 100\% = 20\%$$

$\because v_{\text{Math}} > v_{\text{Chinese}}$

$\therefore$ Mathematics test is more variable than Chinese language test.

### 18.5.2 Exercise 18.5

1. The statistics of the height and weight if grade 1 students in a primary school are as follows:

| | Mean | Standard Deviation |
|---|---|---|
| Height ($cm$) | 115.87 | 4.86 |
| Weight ($cm$) | 19.39 | 2.16 |

Compare the variability of the height and weight of the students.

**Sol.**

$$v_{\text{Height}} = \frac{4.86}{115.87} \times 100\% = 4.19\%$$

$$v_{\text{Weight}} = \frac{2.16}{19.39} \times 100\% = 11.14\%$$

$\because v_{\text{Weight}} > v_{\text{Height}}$

$\therefore$ The weight of the students is more variable than their height.

2. The table below shows the first semester Mathematics exam average mark and standard deviation of five junior 1 classes in a school:

| Class | Average Mark | Standard Deviation |
|---|---|---|
| A | 62 | 11 |
| B | 74 | 9 |
| C | 65 | 10 |
| D | 70 | 7 |
| E | 53 | 8 |

Which class has the smallest coefficient of variation?

**Sol.**

$$v_{\text{A}} = \frac{11}{62} \times 100\% = 17.74\%$$

$$v_{\text{B}} = \frac{9}{74} \times 100\% = 12.16\%$$

$$v_{\text{C}} = \frac{10}{65} \times 100\% = 15.38\%$$

$$v_{\text{D}} = \frac{7}{70} \times 100\% = 10\%$$

$$v_{\text{E}} = \frac{8}{53} \times 100\% = 15.09\%$$

$\because v_D < v_B < v_A < v_C < v_E$

$\therefore$ Class D has the smallest coefficient of variation.

3. The table below shows the Mathematics exam results of two groups of students $A$ and $B$:

| Group | Marks | | | | |
|---|---|---|---|---|---|
| A | 60 | 98 | 76 | 84 | 52 |
| B | 88 | 58 | 90 | 69 | 78 |

(a) Find the average mark of each group.

**Sol.**

$$\bar{x}_A = \frac{60 + 98 + 76 + 84 + 52}{5} = 74$$
$$\bar{x}_B = \frac{88 + 58 + 90 + 69 + 78}{5} = 76.6$$

(b) Find the standard deviation of each group.

**Sol.**

$$\sigma_A^2 = \frac{60^2 + \cdots + 52^2}{5} - 74^2$$
$$= 5748 - 5476$$
$$= 272$$
$$\sigma_A = \sqrt{272} = 16.49$$

$$\sigma_B^2 = \frac{88^2 + \cdots + 78^2}{5} - 76.6^2$$
$$= 6010.6 - 5867.56$$
$$= 143.04$$
$$\sigma_B = \sqrt{143.04} = 11.96$$

(c) Find the coefficient of variation of each group.

**Sol.**

$$v_A = \frac{16.49}{74} \times 100\% = 22.28\%$$
$$v_B = \frac{11.96}{76.6} \times 100\% = 15.61\%$$

4. The table below shows the price of of papayas and grapes per kilogram in the first half of the year (in \$):

| Month | Papaya | Grapes |
|---|---|---|
| January | 3.50 | 20.00 |
| February | 3.00 | 22.00 |
| March | 2.50 | 24.00 |
| April | 3.20 | 23.00 |
| May | 3.60 | 18.00 |
| June | 2.80 | 21.00 |

(a) Find the average price and standard deviation of papayas and grapes respectively in the first half of the year.

**Sol.**

$$\bar{x}_{\text{Papaya}} = \frac{3.50 + \cdots + 2.80}{6} = 3.10$$
$$\bar{x}_{\text{Grapes}} = \frac{20 + \cdots + 21}{6} = 21.33$$

$$\sigma_{\text{Papaya}}^2 = \frac{3.50^2 + \cdots + 2.80^2}{6} - 3.10^2$$
$$= 9.77 - 9.61$$
$$= 0.15$$
$$\sigma_{\text{Papaya}} = \sqrt{0.15} = 0.38$$

$$\sigma_{\text{Grapes}}^2 = \frac{20^2 + \cdots + 21^2}{6} - 21.33^2$$
$$= 459 - 455.11$$
$$= 3.89$$
$$\sigma_{\text{Grapes}} = \sqrt{3.89} = 1.97$$

(b) Which fruit has greater variability in price?

$\because \sigma_{\text{Papaya}} > \sigma_{\text{Grapes}}$

$\therefore$ Papaya has greater variability in price.

5. The table below shows the distribution of annual average marks of two classes of students $A$ and $B$:

| Marks Range | Class $A$ | Class $B$ |
|---|---|---|
| 40 - 49 | 3 | 4 |
| 50 - 59 | 4 | 10 |
| 60 - 69 | 10 | 17 |
| 70 - 79 | 16 | 14 |
| 80 - 89 | 12 | 1 |

Find the coefficient of variation of annual average marks of each class respectively.

**Sol.**

For class $A$,

| Range | $x_i$ | $f_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|---|
| 40 - 49 | 44.5 | 3 | 133.5 | 5940.75 |
| 50 - 59 | 54.5 | 4 | 218 | 11881 |
| 60 - 69 | 64.5 | 10 | 645 | 41602.5 |
| 70 - 79 | 74.5 | 16 | 1192 | 88804 |
| 80 - 89 | 84.5 | 12 | 1014 | 85683 |
| | | $\sum f_i = 45$ | $\sum f_i x_i = 3202.5$ | $\sum f_i x_i^2 = 233911.25$ |

$$\bar{x}_A = \frac{3202.5}{45} = 71.17$$
$$\sigma_A^2 = \frac{233911.25}{45} - 71.17^2$$
$$= 5198.03 - 5064.69$$
$$= 133.34$$
$$\sigma_A = \sqrt{133.34} = 11.55$$
$$v_A = \frac{11.55}{71.17} \times 100\% = 16.23\%$$

For class $B$,

| Range | $x_i$ | $f_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|---|
| 40 - 49 | 44.5 | 4 | 178 | 7921 |
| 50 - 59 | 54.5 | 10 | 545 | 29702.5 |
| 60 - 69 | 64.5 | 17 | 1096.5 | 70724.25 |
| 70 - 79 | 74.5 | 14 | 1043 | 77703.5 |
| 80 - 89 | 84.5 | 1 | 84.5 | 7140.25 |
| | | $\sum f_i = 46$ | $\sum f_i x_i = 2947$ | $\sum f_i x_i^2 = 4193191.5$ |

$$\bar{x}_B = \frac{2947}{46} = 64.07$$
$$\sigma_B^2 = \frac{193191.5}{46} - 64.07^2$$
$$= 4199.82 - 4104.96$$
$$= 95.46$$
$$\sigma_B = \sqrt{95.46} = 9.77$$
$$v_B = \frac{9.77}{64.07} \times 100\% = 15.25\%$$

## 18.6 Correlation and Correlation Coefficient

### Correlation

In statistics, correlation is a statistical measure of the degree to which two or more variables move in relation to each other. For example, the correlation between the height and weight of a person, the correlation between the price of a stock and the volume of the stock traded.

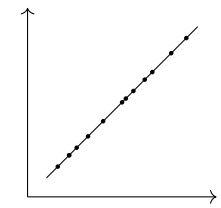
### Scatter Plot

A scatter plot is a type of mathematical diagram to show the relationship between two variables. Let two groups of data be $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$, respectively. The scatter plot of the two groups of data is a graph of the points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$.

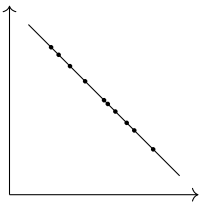
### Linear Correlation

If the scatter plot of two groups of data can be approximated by a straight line, then the two groups of data are said to be linearly correlated. According to the trend of the two groups of data, the correlation can be positive, negative, or zero. For example, the weight of a higher person is usually larger, so the correlation between the weight and height of a person is positive. The sales of a product are usually lower when the price of the product is higher, so the correlation between the price of a product and the volume of the product sold is negative. If there is no relationship between the two groups of data, then it is considered zero correlation.

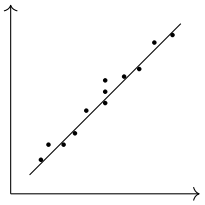
Below are the possible cases of linear correlation:



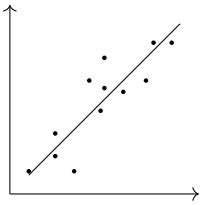
Perfect Positive Correlation



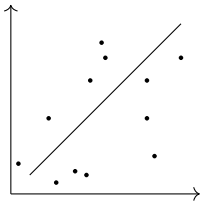
Perfect Negative Correlation



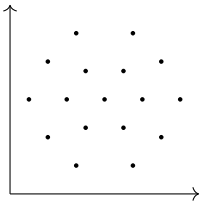
Strong Positive Correlation



Moderate Positive Correlation



Weak Positive Correlation



Zero Correlation

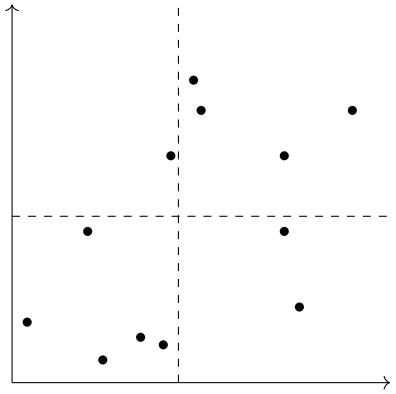
1. If every single point in the scatter plot is on the line of best fit, then it's a perfect positive correlation. If the slope of the line of best fit is positive, then it's a positive correlation. If the slope of the line of best fit is negative, then it's a negative correlation.

2. If the points in the scatter plot are scattered around the line of best fit with non-zero slope, then the closer the points are to the line of best fit, the stronger the correlation is.

3. If the points in the scatter plot are scattered evenly around the whole plot with no obvious pattern, then there is no correlation between the two variables, aka zero correlation.

## Correlation Coefficient

Telling the correlation between two variables by looking at the scatter plot is not a very accurate way. To accurately measure the correlation between two sets of data, we need to use a coefficient that can distinguish the strength of the correlation.



Let the mean value of two sets of data be $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ be $\bar{x}$ and $\bar{y}$ respectively. Draw two lines $x = \bar{x}$ and $y = \bar{y}$ on the scatter plot of the two sets of data, splitting the plot into four quadrants, as shown in the figure above. Now the origin of the plot is at $(\bar{x}, \bar{y})$. If a point $(x_i, y_i)$ is in the first or the third quadrant, then $(x_i - \bar{x})(y_i - \bar{y})$ is positive. As discussed in the previous section, if the correlation is positive, the points are scattering around the line of best fit with positive slope. Therefore, the points are more likely to be in the first or the third quadrant. That means, there are more positive value of $(x_i - \bar{x})(y_i - \bar{y})$ than negative value, therefore the value of $\sum (x_i - \bar{x})(y_i - \bar{y})$ is positive. The higher the correlation is, the more points are in the first or the third quadrant, the higher the positive value of $\sum (x_i - \bar{x})(y_i - \bar{y})$ is.

On the other hand, if a point $(x_i, y_i)$ is in the second or the fourth quadrant, then $(x_i - \bar{x})(y_i - \bar{y})$ is negative, which means there are more negative value of $(x_i - \bar{x})(y_i - \bar{y})$ than positive value, therefore the value of $\sum (x_i - \bar{x})(y_i - \bar{y})$ is negative. Similarly, he higher the correlation is, the lower the negative value of $\sum (x_i - \bar{x})(y_i - \bar{y})$ is.

Hence, the value and the sign of $\sum (x_i - \bar{x})(y_i - \bar{y})$ can be used to measure the correlation between two sets of data. The value of $\sum (x_i - \bar{x})(y_i - \bar{y})$ will be affected by the measurement unit of the data. To make the value of $\sum (x_i - \bar{x})(y_i - \bar{y})$ independent of the measurement unit, we define the correlation coefficient of two sets of data $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ as:

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

The value of $r$ is always between $-1$ and 1. If $r = 0$, then there is no correlation between the two sets of data. If $r > 0$, then the correlation is positive. If $r < 0$, then the correlation is negative. The absolute value of $r$ is the strength of the correlation, and is generally divided as follows:

1. $|r| = 1$: perfect correlation
2. $0 < |r| < 0.3$: weak correlation
3. $0.3 \leq |r| < 0.7$: moderate correlation
4. $0.7 \leq |r| \leq 1$: strong correlation

Dividing both the denominator and the numerator of the formula of $r$ by the number of data points $n$, then the numerator is the mean value of $(x_i - \bar{x})(y_i - \bar{y})$, and the denominator is the product of the standard deviation of $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$. Similar to the standard deviation, there is an easier way to calculate the correlation coefficient:

$$\begin{aligned}
r &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \\
&= \frac{\sum (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})}{\sqrt{\sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \sum (y_i^2 - 2y_i \bar{y} + \bar{y}^2)}} \\
&= \frac{\sum x_i y_i - \bar{y} \sum x_i - y_i \sum \bar{x} + \sum \bar{x}\bar{y}}{\sqrt{\sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \sum (y_i^2 - 2y_i \bar{y} + \bar{y}^2)}} \\
&= \frac{\frac{\sum x_i y_i - \bar{y} \sum x_i - y_i \sum \bar{x} + \sum \bar{x}\bar{y}}{n}}{\frac{\sqrt{\sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \sum (y_i^2 - 2y_i \bar{y} + \bar{y}^2)}}{n}} \\
&= \frac{\frac{\sum x_i y_i}{n} - \frac{\bar{y} \sum x_i}{n} - \frac{y_i \sum \bar{x}}{n} + \frac{\sum \bar{x}\bar{y}}{n}}{\sqrt{\frac{\sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \sum (y_i^2 - 2y_i \bar{y} + \bar{y}^2)}{n \cdot n}}} \\
&= \frac{\frac{\sum x_i y_i}{n} - \frac{\bar{y} \sum x_i}{n} - \frac{\bar{x} \sum y_i}{n} + \frac{n\bar{x}\bar{y}}{n}}{\sqrt{\left(\frac{\sum x_i^2}{n} - \frac{2\bar{x} \sum x_i}{n} + \frac{\sum \bar{x}^2}{n}\right)\left(\frac{\sum y_i^2}{n} - \frac{2\bar{y} \sum y_i}{n} + \frac{\sum \bar{y}^2}{n}\right)}} \\
&= \frac{\frac{\sum x_i y_i}{n} - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y}}{\sqrt{\left(\frac{\sum x_i^2}{n} - 2\bar{x}^2 + \frac{n\bar{x}^2}{n}\right)\left(\frac{\sum y_i^2}{n} - 2\bar{y}^2 + \frac{n\bar{y}^2}{n}\right)}} \\
&= \frac{\frac{\sum x_i y_i}{n} - \bar{x}\bar{y}}{\sqrt{\left(\frac{\sum x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2\right)\left(\frac{\sum y_i^2}{n} - 2\bar{y}^2 + \bar{y}^2\right)}} \\
&= \frac{\frac{\sum x_i y_i}{n} - \bar{x}\bar{y}}{\sqrt{\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right)\left(\frac{\sum y_i^2}{n} - \bar{y}^2\right)}}
\end{aligned}$$

### 18.6.1 Practice 10

1. The table below shows the height (in $cm$) and weight (in $kg$) of 15 10-year-old children:

| Height | Weight |
|---|---|
| 126 | 41 |
| 130 | 42 |
| 110 | 38 |
| 123 | 36 |
| 118 | 33 |
| 130 | 45 |
| 127 | 34 |
| 124 | 35 |
| 116 | 30 |
| 112 | 32 |
| 113 | 31 |
| 121 | 40 |
| 115 | 34 |
| 120 | 35 |
| 118 | 33 |

Calculate the correlation coefficient of the height and the weight of the 15 children, and determine on the strength of the correlation.

**Sol.**

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 126 | 41 | 15876 | 1681 | 5166 |
| 130 | 42 | 16900 | 1764 | 5460 |
| 110 | 38 | 12100 | 1444 | 4180 |
| 123 | 36 | 15129 | 1296 | 4428 |
| 118 | 33 | 13924 | 1089 | 3894 |
| 130 | 45 | 16900 | 2025 | 5850 |
| 127 | 34 | 16129 | 1156 | 4318 |
| 124 | 35 | 15376 | 1225 | 4340 |
| 116 | 30 | 13456 | 900 | 3480 |
| 112 | 32 | 12544 | 1024 | 3584 |
| 113 | 31 | 12769 | 961 | 3503 |
| 121 | 40 | 14641 | 1600 | 4840 |
| 115 | 34 | 13225 | 1156 | 3910 |
| 120 | 35 | 14400 | 1225 | 4200 |
| 118 | 33 | 13924 | 1089 | 3894 |
| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum y_i^2$ | $\sum x_i y_i$ |
| 1803 | 539 | 217293 | 19635 | 65047 |

$$\bar{x} = \frac{1803}{15} = 120.2$$
$$\bar{y} = \frac{539}{15} = 35.93$$
$$r = \frac{\frac{65047}{15} - 120.2 \times 35.93}{\sqrt{\left(\frac{217293}{15} - 120.2^2\right)\left(\frac{19635}{15} - 35.93^2\right)}}$$
$$= 0.6631$$

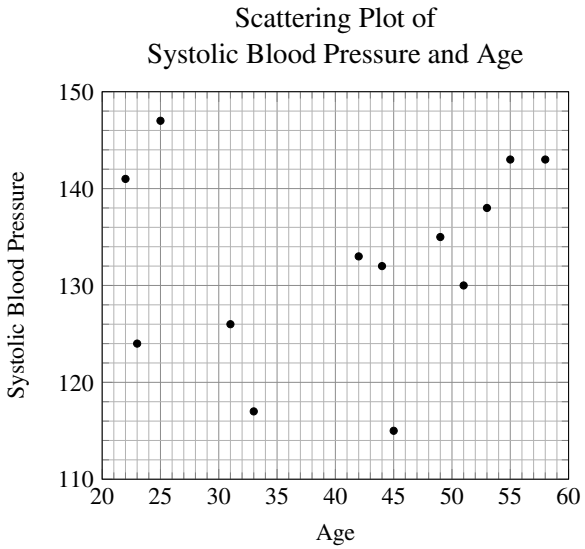
According to the result of the calculation, the height and the weight of the 15 children are positively and moderetely correlated.

2. In order to study the relationship between the systolic blood pressure (in $mmHg$) and the age (in $year$) of human, a medical school collected the data of 13 male patients:

| Age | Systolic Blood Pressure |
|---|---|
| 51 | 130 |
| 22 | 141 |
| 23 | 124 |
| 31 | 126 |
| 33 | 117 |
| 49 | 135 |
| 58 | 143 |
| 53 | 138 |
| 44 | 132 |
| 55 | 143 |
| 42 | 133 |
| 45 | 115 |
| 25 | 147 |

(a) Construct a scatter diagram of the data.

**Sol.**



(b) Calculate the correlation coefficient of the age and the systolic blood pressure of the 13 patients, and determine on the strength of the correlation.

**Sol.**

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 51 | 130 | 2601 | 16900 | 6630 |
| 22 | 141 | 484 | 19881 | 3102 |
| 23 | 124 | 529 | 15376 | 2852 |
| 31 | 126 | 961 | 15876 | 3906 |
| 33 | 117 | 1089 | 13689 | 3861 |
| 49 | 135 | 2401 | 18225 | 6615 |
| 58 | 143 | 3364 | 20449 | 8294 |
| 53 | 138 | 2809 | 19044 | 7314 |
| 44 | 132 | 1936 | 17424 | 5808 |
| 55 | 143 | 3025 | 20449 | 7865 |
| 42 | 133 | 1764 | 17689 | 5586 |
| 45 | 155 | 2025 | 24025 | 6975 |
| 25 | 147 | 625 | 21609 | 3675 |
| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum y_i^2$ | $\sum x_i y_i$ |
| 531 | 1764 | 23613 | 240636 | 72483 |

$$\bar{x} = \frac{531}{13} = 40.85$$

$$\bar{y} = \frac{1764}{13} = 135.69$$

$$r = \frac{\frac{72483}{13} - 40.85 \times 135.69}{\sqrt{\left(\frac{23613}{13} - 40.85^2\right)\left(\frac{240636}{13} - 135.69^2\right)}}$$

$$= 0.2748$$

According to the result of the calculation, the height and the weight of the 15 children are positively and weakly correlated.

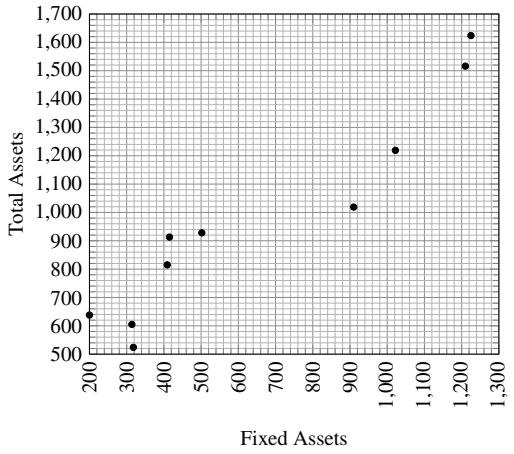
## 18.6.2 Exercise 18.6

1. The table below shows the value of fixed assets and total assets (in 10 thousand dollar) of 10 enterprises of the same industry:

| No. of Enterprise | Fixed Assets | Total Assets |
|---|---|---|
| 1 | 200 | 638 |
| 2 | 314 | 605 |
| 3 | 318 | 524 |
| 4 | 409 | 815 |
| 5 | 415 | 913 |
| 6 | 502 | 928 |
| 7 | 910 | 1019 |
| 8 | 1022 | 1219 |
| 9 | 1210 | 1516 |
| 10 | 1225 | 1624 |

(a) Construct a scatter diagram of the data.

**Sol.**




(b) Calculate the mean value of fixed assets and total assets respectively.

**Sol.**

$$\text{Mean of fixed assets } \bar{x} = \frac{200 + 314 + \cdots + 1225}{10}$$
$$= \frac{6525}{10}$$
$$= 652.5$$
$$= \$652,500,000$$

$$\text{Mean of total assets } \bar{y} = \frac{638 + 605 + \cdots + 1624}{10}$$
$$= \frac{9801}{10}$$
$$= 980.1$$
$$= \$980,100,000$$

(c) Find the correlation coefficient of the fixed assets and the total assets, and determine on the strength of the correlation.

**Sol.**

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 200 | 638 | 40000 | 407044 | 127600 |
| 314 | 605 | 98596 | 366025 | 189970 |
| 318 | 524 | 101124 | 274576 | 166632 |
| 409 | 815 | 167281 | 664225 | 333335 |
| 415 | 913 | 172225 | 833569 | 378895 |
| 502 | 928 | 252004 | 861184 | 465856 |
| 910 | 1019 | 828100 | 1038361 | 927290 |
| 1022 | 1219 | 1044484 | 1485961 | 1245818 |
| 1210 | 1516 | 1464100 | 2298256 | 1834360 |
| 1225 | 1624 | 1500625 | 2637376 | 1989400 |
| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum y_i^2$ | $\sum x_i y_i$ |
| 6525 | 9801 | 5668539 | 10866577 | 7659156 |

$$r = \frac{\frac{7659156}{13} - 652.5 \times 980.1}{\sqrt{\left(\frac{5668539}{13} - 652.5^2\right)\left(\frac{10866577}{13} - 980.1^2\right)}}$$
$$= 0.9478$$

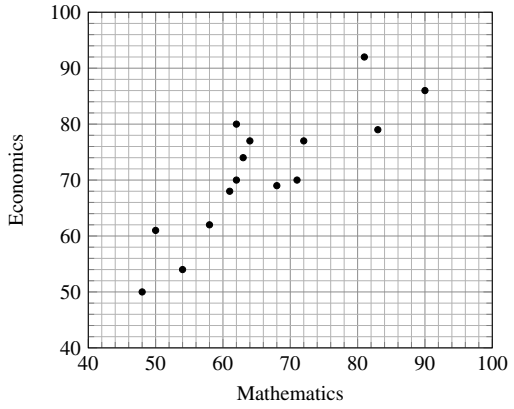
According to the result of the calculation, According to the result of the calculation, the value of fixed assets and total assets of theese 10 enterprises are positively and strongly correlated.

2. The table shows the marks of Mathematics and Ecomoics of 15 students:

| Mathematics | Economics |
|---|---|
| 83 | 79 |
| 50 | 61 |
| 62 | 70 |
| 90 | 86 |
| 68 | 69 |
| 61 | 68 |
| 58 | 62 |
| 62 | 80 |
| 71 | 70 |
| 63 | 74 |
| 72 | 77 |
| 54 | 54 |
| 64 | 77 |
| 48 | 50 |
| 81 | 92 |

(a) Construct a scatter diagram of the data.

**Sol.**




(b) Find the correlation coefficient of the marks of Mathematics and the Economics, and determine on the strength of the correlation.

**Sol.**

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 83 | 79 | 6889 | 6241 | 6557 |
| 50 | 61 | 2500 | 3721 | 3050 |
| 62 | 70 | 3844 | 4900 | 4340 |
| 90 | 86 | 8100 | 7396 | 7740 |
| 68 | 69 | 4624 | 4761 | 4692 |
| 61 | 68 | 3721 | 4624 | 4148 |
| 58 | 62 | 3364 | 3844 | 3596 |
| 62 | 80 | 3844 | 6400 | 4960 |
| 71 | 70 | 5041 | 4900 | 4970 |
| 63 | 74 | 3969 | 5476 | 4662 |
| 72 | 77 | 5184 | 5929 | 5544 |
| 54 | 54 | 2916 | 2916 | 2916 |
| 64 | 77 | 4096 | 5929 | 4928 |
| 48 | 50 | 2304 | 2500 | 2400 |
| 81 | 92 | 6561 | 8464 | 7452 |
| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum y_i^2$ | $\sum x_i y_i$ |
| 987 | 1069 | 66957 | 78001 | 71955 |

$$\bar{x} = \frac{987}{15} = 65.8$$
$$\bar{y} = \frac{1069}{15} = 71.27$$
$$r = \frac{\frac{71955}{15} - 65.8 \times 71.27}{\sqrt{\left(\frac{66957}{15} - 65.8^2\right)\left(\frac{78001}{15} - 71.27^2\right)}}$$
$$= 0.8445$$

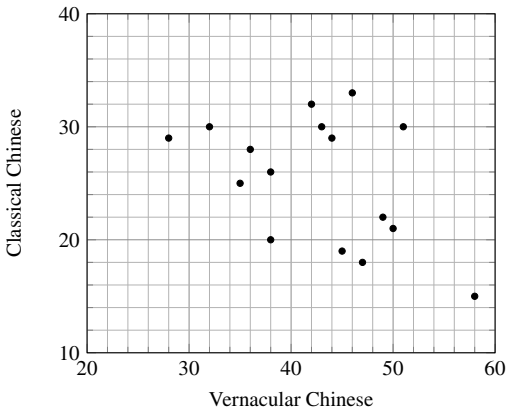
According to the result of the calculation, the marks of Mathematics and the Economics are positively and strongly correlated.

3. The table below shows the marks of 16 students in the Chinese language minor test. The paper was split into two sections: Vernacular and Classical Chinese and their full marks were 60 and 40 respectively.

| Vernacular Chinese | Classical Chinese |
|---|---|
| 43 | 30 |
| 50 | 21 |
| 38 | 20 |
| 45 | 19 |
| 58 | 15 |
| 47 | 18 |
| 32 | 30 |
| 36 | 28 |
| 38 | 26 |
| 51 | 30 |
| 44 | 29 |
| 28 | 29 |
| 49 | 22 |
| 42 | 32 |
| 46 | 33 |
| 35 | 25 |

(a) Construct a scatter diagram of the data.

**Sol.**




(b) Find the correlation coefficient of the marks of Vernacular Chinese and the Classical Chinese, and determine on the strength of the correlation.

**Sol.**

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 43 | 30 | 1849 | 900 | 1290 |
| 50 | 21 | 2500 | 441 | 1050 |
| 38 | 20 | 1444 | 400 | 760 |
| 45 | 19 | 2025 | 361 | 855 |
| 58 | 15 | 3364 | 225 | 870 |
| 47 | 18 | 2209 | 324 | 846 |
| 32 | 30 | 1024 | 900 | 960 |
| 36 | 28 | 1296 | 784 | 1008 |
| 38 | 26 | 1444 | 676 | 988 |
| 51 | 30 | 2601 | 900 | 1530 |
| 44 | 29 | 1936 | 841 | 1276 |
| 28 | 29 | 784 | 841 | 812 |
| 49 | 22 | 2401 | 484 | 1078 |
| 42 | 32 | 1764 | 1024 | 1344 |
| 46 | 33 | 2116 | 1089 | 1518 |
| 35 | 25 | 1225 | 625 | 875 |
| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum y_i^2$ | $\sum x_i y_i$ |
| 682 | 407 | 29982 | 10815 | 17060 |

$$\bar{x} = \frac{682}{16} = 42.62$$

$$\bar{y} = \frac{407}{16} = 25.44$$

$$r = \frac{\frac{17060}{16} - 42.62 \times 25.44}{\sqrt{\left(\frac{29982}{16} - 42.62^2\right)\left(\frac{10815}{16} - 25.44^2\right)}}$$

$$= -0.4444$$

According to the result of the calculation, the marks of Vernacular Chinese and Classical Chinese are negatively and moderately correlated.

4. Below shows the the service costs and values of properties sold by a property broker in 5 trades:

| Service Costs (in \$100) | Value of Prop. (in \$10k) |
|---|---|
| 16.5 | 3.9 |
| 17.4 | 4.2 |
| 16.8 | 4.1 |
| 17.9 | 4.5 |
| 18.4 | 4.8 |

Find the correlation coefficient of the service costs and the values of properties in these 5 trades, and determine on the strength of the correlation.

**Sol.**

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 16.5 | 3.9 | 272.25 | 15.21 | 64.35 |
| 17.4 | 4.2 | 302.76 | 17.64 | 73.08 |
| 16.8 | 4.1 | 282.24 | 16.81 | 68.88 |
| 17.9 | 4.5 | 320.41 | 20.25 | 80.55 |
| 18.4 | 4.8 | 338.56 | 23.04 | 88.32 |
| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum y_i^2$ | $\sum x_i y_i$ |
| 87.0 | 21.5 | 1516.22 | 92.95 | 375.18 |

$$\bar{x} = \frac{87.0}{5} = 17.4$$
$$\bar{y} = \frac{21.5}{5} = 4.3$$
$$r = \frac{\frac{375.18}{5} - 17.4 \times 4.3}{\sqrt{\left(\frac{1516.22}{5} - 17.4^2\right)\left(\frac{92.95}{5} - 4.3^2\right)}}$$
$$= 0.9818$$

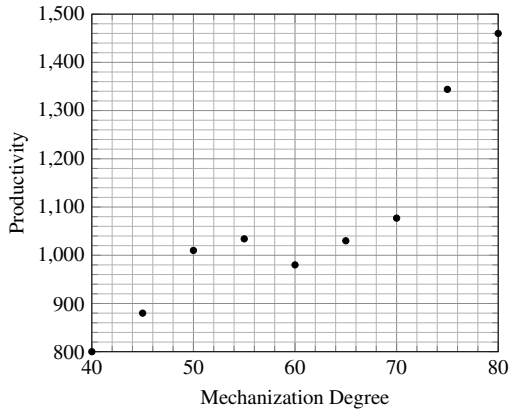
According to the result of the calculation, the service costs and the values of properties are positively and strongly correlated.

5. The table below shows the degree of labor mechanization and labor productivity:

| Mechanization Degree (%) | Productivity ($/pax) |
|---|---|
| 40 | 800 |
| 45 | 880 |
| 50 | 1010 |
| 55 | 1034 |
| 60 | 980 |
| 65 | 1030 |
| 70 | 1077 |
| 75 | 1344 |
| 80 | 1460 |

(a) Construct a scatter diagram of the data.

**Sol.**




(b) Find the correlation coefficient of the degree of labor mechanization and the labor productivity, and determine on the strength of the correlation.

**Sol.**

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 40.0 | 800.0 | 1600.0 | 640000.0 | 32000.0 |
| 45.0 | 880.0 | 2025.0 | 774400.0 | 39600.0 |
| 50.0 | 1010.0 | 2500.0 | 1020100.0 | 50500.0 |
| 55.0 | 1034.0 | 3025.0 | 1069156.0 | 56870.0 |
| 60.0 | 980.0 | 3600.0 | 960400.0 | 58800.0 |
| 65.0 | 1030.0 | 4225.0 | 1060900.0 | 66950.0 |
| 70.0 | 1077.0 | 4900.0 | 1159929.0 | 75390.0 |
| 75.0 | 1344.0 | 5625.0 | 1806336.0 | 100800.0 |
| 80.0 | 1460.0 | 6400.0 | 2131600.0 | 116800.0 |
| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum y_i^2$ | $\sum x_i y_i$ |
| 540.0 | 9615.0 | 33900.0 | 10622821.0 | 597710.0 |

$$\bar{x} = \frac{540.0}{9} = 60.0$$
$$\bar{y} = \frac{9615.0}{9} = 1068.33$$
$$r = \frac{\frac{597710.0}{9} - 60.0 \times 1068.33}{\sqrt{\left(\frac{33900.0}{9} - 60.0^2\right)\left(\frac{10622821.0}{9} - 1068.33^2\right)}}$$
$$= 0.9072$$

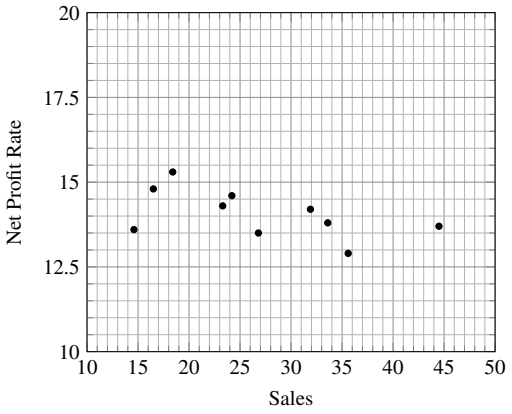
According to the result of the calculation, the degree of labor mechanization and the labor productivity are positively and strongly correlated.

6. Below are the sales (in million) and the net profit rate (%) of 10 department store:

| Company | Sales | Net Profit Rate |
|---|---|---|
| A | 18.4 | 15.3 |
| B | 16.5 | 14.8 |
| C | 14.6 | 13.6 |
| D | 23.3 | 14.3 |
| E | 35.6 | 12.9 |
| F | 24.2 | 14.6 |
| G | 33.6 | 13.8 |
| H | 44.5 | 13.7 |
| I | 26.8 | 13.5 |
| J | 31.9 | 14.2 |

(a) Construct a scatter diagram of the data.

**Sol.**




(b) Find the correlation coefficient of the sales and the net profit rate, and determine on the strength of the correlation.

**Sol.**

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 18.4 | 15.3 | 338.56 | 234.09 | 281.52 |
| 16.5 | 14.8 | 272.25 | 219.04 | 244.2 |
| 14.6 | 13.6 | 213.16 | 184.96 | 198.56 |
| 23.3 | 14.3 | 542.89 | 204.49 | 333.19 |
| 35.6 | 12.9 | 1267.36 | 166.41 | 459.24 |
| 24.2 | 14.6 | 585.64 | 213.16 | 353.32 |
| 33.6 | 13.8 | 1128.96 | 190.44 | 463.68 |
| 44.5 | 13.7 | 1980.25 | 187.69 | 609.65 |
| 26.8 | 13.5 | 718.24 | 182.25 | 361.8 |
| 31.9 | 14.2 | 1017.61 | 201.64 | 452.98 |
| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum y_i^2$ | $\sum x_i y_i$ |
| 269.4 | 140.7 | 8064.92 | 1984.17 | 3758.14 |

$$\bar{x} = \frac{269.4}{10} = 26.94$$
$$\bar{y} = \frac{140.7}{10} = 14.07$$
$$r = \frac{\frac{3758.14}{10} - 26.94 \times 14.07}{\sqrt{\left(\frac{8064.92}{10} - 26.94^2\right)\left(\frac{1984.17}{10} - 14.07^2\right)}}$$
$$= -0.5350$$

According to the result of the calculation, the sales and the net profit rate are negatively and moderately correlated.

# 18.7 Statistical Index

## Index

In statistics, an index is a number that measures the changes in a figure from one point in time to another. There is a wide range of applications of an index, such as the price index which represents the changes in prices, the production index which represents the changes in production, and the wage index which represents the changes in salaries and wages. There are also other indices such as the living index, foreign exchange index, population index, stock market index, etc.

The index is a kind of relative number. The standard period that is used for comparison when calculating the index is called the base period. As the case may be, the base period can be a year or a month. The index of the base period is usually a number that is easier to be remembered and compared, such as 100, 500, or 1000, and the chosen number must be able to represent the changes in the figure. We will use 100 as our base period index. The period that is used for comparison to the base period is called the current period. Let $Q_0$ be the base period index and $Q_1$ be the current period index. The index of the current period is calculated by the following formula:

$$I = \frac{Q_1}{Q_0} \times 100$$

where 100 is the base period index.

## Price Relative

The price relative is a simple index that compares the prices of products in different periods. Let $P_0$ be the price of a product in the base period and $P_1$ be the price of the same product in the current period. The price relative of the current period is calculated by the following formula:

$$I = \frac{P_1}{P_0} \times 100$$

### 18.7.1 Practice 11

The table below shows the net profits (in million) of a company from 2010 to 2014. Use the year 2010 as the base period and calculate the index of the net profit of the company in each year.

| Year | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|
| Net Profit | 700 | 621 | 584.1 | 720.5 | 800 |

### 18.7.2 Exercise 18.7a

1. The prices of white sugar in 2011, 2012, and 2013 are \$2.10, \$2.30, and \$2.50 respectively. Use the year 2011 and 2012 as the base period and calculate the price relative of the year 2013.

2. The prices of a food product in 2011, 2013, and 2015 are \$3.40, \$3.75, and \$3.90 respectively. Use the year 2011 as the base period and calculate the price relative of the year 2013 and 2015.

3. The number of new students of a school from 2011 to 2015 are as follows:

| Year | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|
| New Stud. | 182 | 150 | 120 | 104 | 94 |

   Use the year 2011 as the base period and calculate the index of the number of new students in each year.

4. The table below shows the price of terraced houses (in \$10k) of a place in from 2019 to 2014:

| Year | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|
| Price | 32.0 | 35.5 | 43.4 | 51.0 | 60.0 |

   Use the year 2010 as the base period and calculate the index of the price of terraced houses in each year.

5. The table below shows the price relative of three products $A$, $B$, and $C$ when using different years as the base period and the current period:

| Current Period | Base Period | $A$ | $B$ | $C$ |
|---|---|---|---|---|
| 2010 | 2005 | 160 | $x$ | 170 |
| 2015 | 2005 | 140 | 190 | $y$ |
| 2015 | 2010 | $z$ | 210 | 150 |

   Find the value of $x$, $y$, and $z$.

## Composite Index

The composite index is the mean value of indices of different figures. Since the importance of each figure might be different, the weight of each index is used to represent the importance of each figure, and the acquired weighted mean is called the composite index.

Let the simple index of $n$ figures of the same base period and the same current period be $x_1, x_2, \ldots, x_n$, and their respective weights be $w_1, w_2, \ldots, w_n$. The composite index is calculated by the following formula:

$$\begin{aligned}\bar{I} &= \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n} \\ &= \frac{\sum w_ix_i}{\sum w_i}\end{aligned}$$

If the study object is some product, where $x_i$ is the price relative to the $i^{th}$ product, then its weighted mean is called the price index. If the study object is the daily living expenses, then its weighted mean is called the living consumption index.

### 18.7.3 Practice 12

The table below shows the prices and weights of sneakers of three brands in 2012 and 2015:

| Sneakers | Unit Price | | Weight |
|---|---|---|---|
| | 2012 | 2015 | |
| A | 230 | 233 | 5 |
| B | 225 | 228 | 3 |
| C | 215 | 221 | 2 |

1. Use the year 2012 as the base period and calculate the price relative of each brand in 2015.

2. Use the year 2012 as the base period and calculate the price index of sneakers in 2015.

### 18.7.4 Exercise 18.7b

1. Using 2012 as the base period, the price relatives of foods, gases and clothes in 2014 are 111, 105, and 106 respectively, and their weights are 5, 1, and 2 respectively. Calculate the composite index of the three comsumer items in 2014.

2. The table below shows the price of each primary food in 2015 (with 2005 as the base period). Find the price index in 2015.

| Food | Price Relative | Weight |
|---|---|---|
| Meat | 130 | 15 |
| Fish | 150 | 14 |
| Vegetable | 200 | 10 |
| Rice | 110 | 20 |
| Cooking Oil | 120 | 8 |
| Beverage | 150 | 7 |
| Fruit | 160 | 6 |

3. The weight and unit price of 3 kind of materials bought by a factory are as follows:

| Material | Weight (ton) | Unit Price ($) | |
|---|---|---|---|
| | | 2010 | 2014 |
| A | 20 | 0.62 | 0.71 |
| B | 50 | 2.05 | 2.09 |
| C | 60 | 0.80 | 0.85 |

Using 2010 as the base period, 2014 as the current period,

(a) Find the composite index of the unit prices of the three materials without considering the weights (i.e. the weights are all 1).

(b) Using the weight of each material as the weight, find the composite index of the unit prices of the three materials.

4. The table below shows three indices and their weights. If their composite index is 103, find the value of $x$.

| Index | 90 | $11x$ | 120 |
|---|---|---|---|
| Weight | $x$ | 4 | 6 |

5. The table below shows the price relative and weight of three products with 2013 as the base period and 2015 as the current period. Given that the price of item $A$ in 2013 and 2015 are \$20 and \$25 respectively, the price of item $B$ is twice the price of item $A$.

| Item | Price Relative | Weight |
|---|---|---|
| A | $r$ | 2 |
| B | $t$ | 1 |
| C | 120 | 3 |

(a) Find the value of $r$ adn $t$.

(b) Using 2013 as the base period, find the price index in 2015.

6. The table below shows the price relative and weight of 5 products with 2012 as the base period and 2014 as the current period:

| Item | Price Relative | Weight |
|---|---|---|
| A | 125 | 2 |
| B | 120 | $3x$ |
| C | 110 | 2 |
| D | 130 | $x$ |
| E | 115 | 2 |

Given that the price index in 2014 is 120,

(a) Find the value of $x$.

(b) Assume that the price of item $A$ in 2014 is \$30, find the price of the item in 2012.

7. The table below shows the price, price relative and weight of 4 products in 2012 and 2014:

| Item | Price ($) | | Price Relative | Weight |
|---|---|---|---|---|
| | 2012 | 2014 | | |
| A | 12 | $y$ | 150 | 1 |
| B | $x$ | 24 | 120 | 2 |
| C | 14 | 28 | $z$ | 3 |
| D | 10 | 13 | 130 | 4 |

where the base period of the price relative is 2012, and the current period is 2014.

(a) Find the value of $x$, $y$ and $z$.

(b) Using 2012 as the base period, find the price index in 2014.

8. The table below shows the price of two products in 2005 and 2015:

| Item | Price ($) | | Price Relative |
|---|---|---|---|
| | 2005 | 2015 | |
| A | 30 | $x$ | 2 |
| B | 50 | $x+10$ | 3 |

## 18.8 Revision Exercise 18

1. The length of 60 cotten fibers (in $mm$) in a laboratory are as follows:

82 202 352 321 25 293 293 86
28 206 323 355 357 33 325 113
233 294 50 296 115 236 357 326
52 301 140 328 238 358 58 255
143 360 340 302 370 343 260 303
59 146 60 263 170 175 348 305
380 346 61 305 264 383 62 306
195 350 265 385

(a) Use $21mm$ as the lower limit and $40mm$ as the class range, construct a frequency distribution table.

(b) Construct a histogram and a frequency polygon.

(c) Construct a cumulative frequency table and a cumulative frequency polygon.

(d) Using the cumulative frequency polygon, find the percentage of fibers whose length is greater than $150mm$.

(e) Find the interquartile range.

2. Find the mean, median, range, quartile deviation, and mean deviation of the data $8, 10, 9, 12, 4, 4, 2$.

3. The weight (in $kg$) if 16 babies are as follows:

$$\begin{array}{cccccccc} 8 & 9 & 10 & 9 & 8 & 7 & 9 & 10 \\ 9 & 8 & 8 & 9 & 10 & 9 & 8 & 7 \end{array}$$

Find the mean, meidan, mode, range, quartile deviation, mean deviation, and standard deviation of their weights.

4. The table below shows the score distribution of business study minor test of senior 3 students in a high school:

| Marks | No. of Students |
|---|---|
| 0-9 | 7 |
| 10-19 | 21 |
| 20-29 | 32 |
| 30-39 | 27 |
| 40-49 | 13 |

(a) Construct a cumulative frequency distribution table.

(b) Construct a cumulative frequency polygon.

(c) Find the median and the interquartile range from the cumulative frequency polygon.

(d) Find the percentage of students who scored higher or equal to 45 marks.

(e) Assume that the passing score is 15 marks. Find the percentage of students who failed the test.

5. The burning time (in $s$) of 10 rocket boosters are as follows:

$$\begin{array}{ccccc} 50.7 & 54.9 & 54.3 & 44.8 & 42.2 \\ 69.0 & 55.4 & 66.1 & 48.1 & 34.5 \end{array}$$

Find the range, variance and standard deviation of the burning time.

6. The table below shows the scores of 30 rounds of game scored by someone:

| Score | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Times | 5 | 3 | 4 | $x+1$ | 7 |

Find:

(a) The value of $x$.

(b) The mean and standard deviation of the scores.

7. The table below shows the distribution of scores of a minor test of students in a class:

| Score | No. of Students |
|---|---|
| $0 < x \leq 5$ | 8 |
| $5 < x \leq 10$ | 1 |
| $10 < x \leq 15$ | 9 |
| $15 < x \leq 20$ | 7 |
| $20 < x \leq 25$ | 11 |
| $25 < x \leq 30$ | 4 |

Find:

(a) Range

(b) Median

(c) Mode

8. Below are the distribution of scores of business study exam of 40 students in a class:

| Score | No. of Students |
|---|---|
| 46 - 54 | 4 |
| 54 - 62 | 9 |
| 62 - 70 | 10 |
| 70 - 78 | 8 |
| 78 - 86 | 6 |
| 86 - 94 | 3 |

Find:

(a) Mean

(b) Median

(c) Mode

(d) Variance

9. The table below shows the frequency distribution of the life of 500 light bulbs:

| Life (in $hr$) | No. of Bulbs |
|---|---|
| 800 - 850 | 35 |
| 850 - 900 | 127 |
| 900 - 950 | 185 |
| 950 - 1000 | 103 |
| 1000 - 1050 | 42 |
| 1050 - 1100 | 8 |

Find:

(a) The mean and standard deviation of the life of the light bulbs.

(b) Mean deviation.

(c) Median.

(d) Quartile deviation.

10. Assume that the mean value of data $2, x+1, 5, 2x+1, 8, 2x-3$ is 4,

(a) Find the value of $x$.

(b) With that, find the standard deviation of the data.

11. The mean and mode of a set of data $2, 5, 3, 11, 9, 2, 11, p, q$ are 6 and 3 respectively, $p > q$. Find

(a) The value of $p$ and $q$

(b) Median

(c) Standard deviation

12. Given that the mean value of $x, x+1, 2x-3, 5, y, 8$ is 6. After eliminating $y$, the mean value of the remaining data is 3.8.

(a) Find the value of $x$ and $y$.

(b) With that, find the variance of the original 6 data.

13. Given the sum of the square of 10 numbers is 400, and their mean value is 5. If a number 8 is eliminated from the data set, find the mean value and variance of the remaining data.

14. There are two female chorus groups $A$ and $B$, each of which has 5 members. Their heights (in $cm$) are as follows:

| | | | | | |
|---|---|---|---|---|---|
| Group A | 170 | 162 | 159 | 160 | 155 |
| Group B | 180 | 165 | 150 | 154 | 160 |

(a) Find the mean and standard deviation of the heights of the members of the two groups.

(b) Which group has a lower height variance?

15. The table below shows scores of maths exam of three classes:

| Class | Avg. Marks | Std. Deviation | No. of Stud. |
|---|---|---|---|
| A | 36.8 | 5.2 | 32 |
| B | 30.3 | 12.4 | 36 |
| C | 38.8 | 10.3 | 32 |

(a) In between class $A$, $B$ and $C$, which class has the most consistent performance? Why?

(b) Find the average marks and standard deviation of these three classes combined.

16. The score given by six judges to a gymnast are as follows:

7 5 9 7 8 6

Find the following of the gymnast:

(a) Mean

(b) Standard deviation

(c) Correlation Coefficient

17. In an IQ test, the average score of 10 students is 114, and the scores of 9 of them are as follows:

101 125 118 118 128 106
115 99 118 109

Find:

(a) The IQ of the 10th student.

(b) The correlation coefficient of the IQ of the 10 students.

18. Given that the data of the weight of two groups of girls (in $kg$) are as follows:

| | Mean | Std. Dev. |
|---|---|---|
| 1 years old | 10.90 | 1.24 |
| 5 years old | 19.00 | 2.11 |

Compare the strength of correlation of the weight of these girls.

19. The prodution output and production cost of a factory in the first half of this year are as follows:

| Month | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Output (in $1k$ *tons*) | 2 | 3 | 1 | 4 | 3 | 5 |
| Cost (in $\$1k$) | 9 | 11 | 7 | 13 | 11 | 15 |

20. The marks of Chinese exam and Maths exam of 16 senior students in a school are as follows:

| Chinese | Maths |
|---|---|
| 82 | 59 |
| 79 | 63 |
| 76 | 99 |
| 63 | 67 |
| 56 | 61 |
| 67 | 82 |
| 69 | 82 |
| 81 | 77 |
| 77 | 75 |
| 73 | 74 |
| 58 | 67 |
| 64 | 79 |
| 68 | 75 |
| 72 | 65 |
| 75 | 64 |
| 80 | 66 |
| 83 | 68 |

(a) Construct a scatter diagram of the data.

(b) Find the correlation coefficient of the two exams, and determine the strength of correlation.

21. The table below shows the prices of a product (in $) in 2005, 2010, and 2015:

| Year | 2005 | 2010 | 2015 |
|---|---|---|---|
| Price | 4 | 6 | $x$ |

(a) Assume that the percentage of price increase from 2005 to 2010 is the same as that from 2010 to 2015, find the value of $x$.

(b) Find the price relative in 2015 with respect to 2005.

22. The price data of primary food of a city with 2013 as base period and 2014 as current period are as follows:

| Food | Price Relative | Weight |
|---|---|---|
| Meat | 105 | 8 |
| Fish | 111 | 7 |
| Vegetables | 98 | 5 |
| Rice & Noodles | 103 | 10 |
| Cooking Oil | 100 | 3 |
| Beverage | 107 | 2 |
| Fruits | 99 | 2 |

Find the price index in 2014.

23. The price relative of daily expenses of people in a place with repsect to last year and their relative consumption are as follows:

| Daily Expenses | Price Relative | Consumption Relative |
|---|---|---|
| Clothing | 120 | 23 |
| Food | 117 | 40 |
| Housing | 132 | 19 |
| Transportation | 130 | 18 |

Using the relative consumption as weight, find the composite price index of daily expenses.

24. The table below shows the spending of a company in 4 different projects in 3 consecutive years:

| Project | Year | | | A | B |
|---|---|---|---|---|---|
| | 2012 | 2013 | 2014 | | |
| Salaries | $x$ | 20,000 | 30,000 | 150 | $P$ |
| Stationery | 5,000 | $y$ | 7,000 | 120 | 140 |
| Repair | 4,000 | 5000 | $z$ | 125 | 150 |
| Miscellaneous | 8,000 | $Q$ | 15,000 | $R$ | $R$ |

Given that $A$ is the index where 2012 is the base period and 2013 is the current period; $B$ is the index where 2012 is the base period and 2014 is the current period. Find the value of $x$, $y$, $z$, $P$, $Q$, and $R$.

# Chapter 19

# Permutations and Combinations

Permutations and combinations are the foundation of probability and statistics. In our daily life, we often need to calculate the number of ways of completing a task. These calculations are based on two basic principles: addition principle and multiplication principle.

## 19.1 Addition and Multiplication Principles

**Theorem 1.** *Addition Principle*

*If there are n methods of doing a task, the first method can be done in $m_1$ ways, the second method can be done in $m_2$ ways, $\cdots$, the nth methods can be done in $m_n$ ways, and they are mutually exclusive, which means he task can be done in whatever way using whatever method, then the total number of ways of doing the task is*

$$m_1 + m_2 + \cdots + m_n$$

**Theorem 2.** *Multiplication Principle*

*If there are n steps in doing a task, the first step can be done in $m_1$ ways, the second step can be done in $m_2$ ways, $\cdots$, the nth steps can be done in $m_n$ ways, then the total number of ways of doing the task is*

$$m_1 \times m_2 \times \cdots \times m_n$$

### 19.1.1 Practice 1

1. There are 2 Math reference books, 3 novels, and 4 storybooks of idioms. Xiao Hua wants to choose one book from each category. How many ways can he choose?
2. Travelling from $A$ to $B$ can be done by bus or train. There are 4 buses and 3 trains. How many ways are there to travel from $A$ to $B$?

### 19.1.2 Practice 19.1

1. During the eve of a festival, there are 3 trains, 4 buses, and 4 trains from Johor Bahru to Pinang. How many ways are there to travel from Johor Bahru to Pinang during the day?
2. One has 5 shirts and 6 pants, how many ways can he dress up?
3. There are 4 airlines $A$, $B$, $C$, and $D$ that provide flights from Kuala Lumpur to Bangkok: $A$ provdies 3 flights per day, $B$ provides 2 flights per day, $C$ and $D$ provides 1 flight per day. How many choices are there to travel from Kuala Lumpur to Bangkok?
4. How many set meal combinations are there if there are 6 type of main dishes, 5 type of drinks, and 2 type of desserts?
5. There are 4 doors in a classroom, student $A$ and student $B$ can enter the classroom through any door. How many ways are there for student $A$ and student $B$ to enter the classroom?
6. A friendly match is held between 2 ping pong teams, each team has to send 3 players, and each player has to play games with all the other players on the other team. How many games have to be played?
7. Matching 8 clothes of different colors with 5 different skirts, how many ways are there to dress up? If the above dresses are paired with 4 pairs of shoes of different colors, how many ways are there to dress up?

## 19.2 Permutations and Permutation Formula

### 19.2.1 Practice 2

How many ways are there to arrange the numbers 1, 2, 3, 4 into a two digit number with no repeated digits?

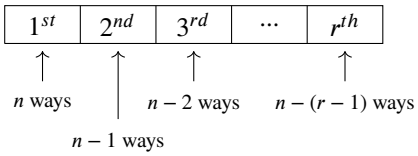
**Sol.**

First step: choose one of the 4 numbers as the first digit, there are 4 ways to do so.

Second step: choose one of the remaining 3 numbers as the second digit, there are 3 ways to do so.

According to the multiplication principle, the total number of ways to arrange the numbers 1, 2, 3, 4 into a two digit number with no repeated digits is $4 \times 3 = 12$.

If there are $n$ elements, we want to pick $r$ elements from them and arrange them in a sequence, how many ways are there to do so? This question can be treated as there are $r$ empty boxes, which means this requires $r$ steps to complete.




First step: CHhoose one element from $n$ elements and put it in the first box, then there are $n$ ways to do so.

Second step: Choose one element from $n-1$ elements and put it in the second box, then there are $n-1$ ways to do so.

Third step: Choose one element from $n-2$ elements and put it in the third box, then there are $n-2$ ways to do so.

So on and so forth, when $r-1$ boxes are filled, the last box $r$ can only be filled with one of the remaining $n-(r-1)$ elements, so there are $n-(r-1)$ ways to do so. According to the multiplication principle, the total number of ways to fill in $r$ boxes is

$$n(n-1)(n-2)\cdots(n-r+1)$$

Threrefore, there are $n(n-1)(n-2)\cdots(n-r+1)$ ways to arrange $r$ elements, and this denoted as ${}^nP_r$, ${}_nP_r$, or $P_r^n$.

$${}_nP_r = n(n-1)(n-2)\cdots(n-r+1)$$

Where $r \leq n$, $n \in N$, $r = 0, 1, 2, \cdots, n$. This formula is called the permutation formula.

When $r = n$, aka a full permutation, the formula becomes

$${}_nP_n = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$$

Therefore, the permutation of all $n$ elements is equal to the products of natural numbers from 1 to $n$. This is called the factorial of $n$, denoted as $n!$.

$$\begin{aligned} n! &= {}_nP_n \\ &= n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 \end{aligned}$$

Using factorial, the permutation formula can be transform into the following:

$$\begin{aligned} {}_nP_r &= n(n-1)(n-2)\cdots(n-r+1) \\ &= \frac{n(n-1)(n-2)\cdots(n-r+1)(n-r)\cdots 3\cdot 2\cdot 1}{(n-r)\cdots 3\cdot 2\cdot 1} \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

Hence, the permutation formula can be written as

$${}_nP_r = \frac{n!}{(n-r)!}$$

Note: 0! is defined as 1 to make the formula work when $n = r$.

### 19.2.2 Practice 3

1. Find the value of ${}_7P_3$ and 5!.
2. Calculate ${}_{10}P_3 + {}_8P_4$.
3. If $100({}_nP_2) = {}_{2n}P_3$, find the value of $n$.

### 19.2.3 Exercise 19.2a

1. Write down all the permutations of 3 elements in 4 elements $A$, $B$, $C$, $D$.
2. Calculate:
   (a) ${}_{15}P_4$
   (b) ${}_{100}P_3$
   (c) 7!
   (d) $\frac{8!}{5!}$
3. Calculate the following:
   (a) $\frac{11!-10!}{10!-9!}$
   (b) $\frac{7!-6!-5!}{5!}$
   (c) $\frac{13!-12!}{(12)^2 10!}$
   (d) $\frac{5({}_8P_3)}{2({}_6P_2)}$
   (e) $\frac{{}_9P_5+{}_9P_4}{{}_9P_3}$
   (f) $\frac{{}_nP_{12}12-{}_nP_{12}11}{{}_nP_{10}10}$
4. Simplify the following:
   (a) $\frac{(n+1)!}{(n-1)!}$
   (b) $\frac{(20-r)!}{(18-r)!}$
5. Find the value of $n$ or $r$ of the following expressions:
   (a) $\frac{(n+1)!}{n!} = 42$

(b) $127({}_{2n}P_3) = {}_{2n}P_4$

(c) $18({}_{n-1}P_2) = {}_{n}P_4$

(d) ${}_{2n+1}P_4 = 132({}_{n}P_n3)$

(e) $4({}_{10}P_{r-1}) = {}_{10}P_r$

(f) $6({}_{9}P_{r-2}) = {}_{9}P_r$

### 19.2.4 Practice 4

1. How many 3 digit numbers with no repeated digits can be formed using the digits 1, 2, 3, 4, 5?

2. How many 3 digit numbers with no repeated digits can be formed using the digits 0, 1, 2, 3, 4?

### 19.2.5 Practice 5

1. There are 50 seats and 50 students in a class. How many ways can the students be seated in the class?

2. Person *A* and *B* has two choose two adjacent seats in a row of 5 chairs. How many ways can they be seated?

3. 4 boys and 2 girls are standing in a row to take a photo. Assume that the two girls has to stand next to each other, how many ways can they be arranged?

### 19.2.6 Exercise 19.2b

1. Assume that there is no repeated digits, how many 5 digit numbers can be formed using the digits 1, 2, 3, 4, 5?

2. How many ways are there to arrange the flags of 10 ASEAN members in a row?

3. 7 novel stories are to be compiled into a book. The sortest story must be placed at the beginning of the book, while the longest story must be placed at the end. How many ways can the stories be arranged?

4. Ten students are to be arranged in a row. Two of the tallest students must be placed at the beginning of the row. How many ways can the students be arranged?

5. There are nine programmes in a literature festival. If one of the programmes is to be placed at the middle or at the end, how many ways can the programmes be arranged?

6. How many permutations of the letters in the word *EQUATION* are there? if the letter *E* and *N* are to be placed at the beginning and at the end respectively, how many ways can the letters be arranged?

7. There are 4 mobile phones that are to be registered a mobile phone number. Chosen 7 phone numbers for pairing, how many ways can the mobile phones be paired with the phone numbers?

8. There are 4 passengers sitting inside a 6 seats SUV. How many ways can the passengers be seated in the SUV?

9. A ping pong coach wants to choose 3 players from a total of 5 players to be the first single, second single and third single respectively. If an elite player has to be chosen as the first or the second single, how many ways can the players be chosen?

10. 8 chairs are to be arranged in two rows of 4 chairs each, in order to provide a place for 8 people to sit. If 3 out of the 8 people are to be seated in the first row, how many ways can the chairs be arranged?

11. How many permutations of the 5 different letters in the word *TRIANGLE* are there? If the beginning and the end of the word are consonants, how many ways can the letters be arranged?

12. All the letters in the word *FANCIES* are to be arranged. If the vowels are to be arranged at even positions, how many ways can the letters be arranged?

13. Rearranging all the letters in the word *NUMERIACAL*, how many ways can the letters be arranged? if all vowels are to be put together, how many ways can the letters be arranged?

14. Examinations of 7 subjects are to be arranged in a row of 7 days, with one subject at a day. If two of which cannot be arranged to be held on two consecutive days, how many ways can the examinations of these 7 subjects be arranged?

15. From 5 numbers 1, 2, 3, 4, 5, how many ways can the following numbers with no repeated digits be formed:

    (a) 5 digits odd numbers

    (b) 5 digits even numbers

16. If the digits are not repeated, from the numbers 0 to 5, how many ways can the 6 digits odd numbers be formed?

17. From the 8 numbers 0, 1, 2, 3, 4, 5, 6, 7, how many 5 digits numbers can be formed such that their digits are not repeated and can be divided by 25?

18. From the 6 numbers 0, 1, 2, 3, 4, 5, how many 4 digits numbers can be formed such that their digits are not repeated and can be divided by 4?

## 19.3 Circular Permutations

In the permutation we have discussed in the previous section, all the elements are arranged in a row. This kind of permutation is called *linear permutation*. Its identity is that it has a beginning and an end. The permutations we are going to discuss in this section are arranged on a closed curve line. This kind of permutation is called *circular permutation*. The identity of circular permutation is that it has no beginning and no end.

For this, we will use an example to explain the concept of circular permutation.

Four people are to be seated in a circle. How many ways can they be seated?

**Sol.**

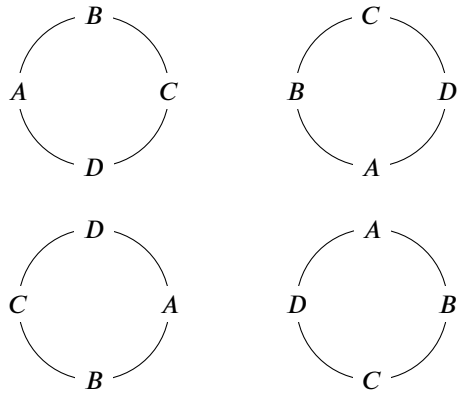
Let these four people be $A$, $B$, $C$ and $D$ respectively.

If these people are arranged in a row, there are 4! ways to arrange them.

Let's take a look at the arrangements below:

$$\begin{array}{cccc} A & B & C & D \\ B & C & D & A \\ C & D & A & B \\ D & A & B & C \end{array}$$

You'll notice that the arrangements above are all the same if being arranged on a circle, and the only difference is the position of the first person.




From the above example, each circular permutation is corresponded to 4 linear permutations. Given that the permutation of 4 people arranging in a row has 4! ways, so there are $\frac{4!}{4} = 3! = 6$ ways to arrange these four people in a circle.

If we generalize the above example, we can get the following formula:

1. The formula of circular permutation of $n$ elements:

$$\frac{{}_nP_n}{n} = \frac{n!}{n} = (n-1)!$$

2. The formula of circular permutation of $r$ elements from $n$ elements ($r \leq n$):

$$\frac{{}_nP_r}{r} = \frac{n!}{r(n-r)!}$$

### 19.3.1 Practice 6

1. Choose 5 people from 6 males and 5 females to be seated in a circle. How many ways can they be seated?

2. 6 males and 4 females are to be seated around a circular table. If females can't seat beside each other, how many ways can they be seated?

### 19.3.2 Exercise 19.3

1. 8 people are to be seated in a circle. How many ways can they be seated?

2. 10 children are to be arranged in a circle. How many ways can they be arraged? If a child must be seated in the primary position, how many ways can they be arranged?

3. 6 people are to be formed into a circle. If two people must seat together, how many ways can they be arranged?

4. 4 males and 3 females are to be seated around a circular table. If none of the females can seat together, how many ways can they be seated?

5. 4 pairs of couples and one child are to be seated around a circular table. If the couples must sit together, how many ways can they be seated?

6. A family of 7 people are sitting together around a circular table for a dinner. If the grandfather, grandmother, father and mother must sit together, of which the grandfather and the grandmother, the father and the mother must sit together, how many ways can they be seated?

7. If a linear permutation of $n$ people is 6 times of a circular permutation of $n$ people, find these two permutations.

## 19.4 Full Permutations of Inexactly Distinct Elements

In all the previous questions of permutations, the given elements are all distinct. However, in some cases, there are some elements that are the same, this kind of permutation are considered as *permutations with repetition*. Let's discusse the folowing example:

For a full permutation of three elements $a$, $a$, and $b$, how many ways can they be arranged?

Let's treat the identical elemnents $a$ as two different elements $a_1$ and $a_2$, there will be 3! ways to arrange these three different element $a_1$, $a_2$, and $b$, as listed below:

$$\begin{array}{cccccc} a_1 & a_2 & b & & a_2 & a_1 & b \\ a_1 & b & a_2 & & a_2 & b & a_1 \\ b & a_1 & a_2 & & b & a_2 & a_1 \end{array}$$

In the three rows above, for each row, the position of $b$ is fixed, $a_1$ and $a_2$ has $2! = 2$ ways to be arranged.

If we change $a_1$ and $a_2$ above back to $a$, then the two different arrangements of $a_1$ and $a_2$ will be counted as one arrangement of $a$. Hence, there are only $\frac{3!}{2!} = 3$ ways for full permutation of three elements $a$, $a$, and $b$.

Generalize the above example, given $n$ elements, where there are $n_1$ elements $a_1$, $n_2$ elements $a_2$, $\cdots$, $n_k$ elements $a_k$, where $n_1 + n_2 + \cdots + n_k = n$, then the number of full permutations of these $n$ elements is:

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

### 19.4.1 Practice 7

1. Giving 9 children 2 pens, 3 ball-point pens, 4 pencils, how many ways can they be given if each child must be given one pen?

2. Find the full permutation of the letters in the word *EXPESSION*.

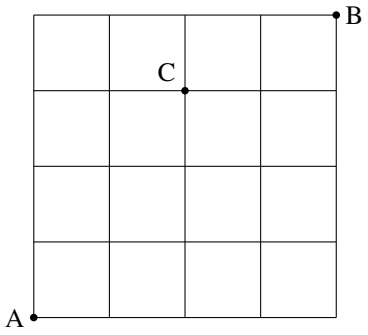
### 19.4.2 Exercise 19.4

1. How many 8 digit numbers can be formed using 8 digits 1, 1, 2, 2, 3, 3, 4, 5?

2. Arranging all the letters in the word *MALAYSIA*, how many ways can they be arranged such that the three "A"s are not totally together?

3. Arranging all the letters in the word *MATHEMATICAL*, how many ways can they be arranged such that the three "A"s are not totally together?

4. The diagram below shows a city with 5 north-west roads and 5 south-east roads, where each road shares the same length.




   (a) How many shortest paths are there from A through C to B?

   (b) How many shortest paths are there from A to B without passing through C?

5. Arranging all the letters in the word *GEOMETRIC*, find the number of permutations such that the following conditions are satisfied:

   (a) No limitaion on the arrangement of the letters.

   (b) All the vowels must be together.

   (c) None of the vowels are adjacent to each other.

## 19.5 Permutations with Repetition

If the elements can be chosen more than once, arranging $r$ elements from $n$ elements is called *permutation with repetition* of $r$ elements from $n$ elements.

Given $n$ distinct elements, each element can be chosen more than once, then there are $n$ ways to choose the first element, $n$ ways to choose the second element, $\cdots$, $n$ ways to choose the $r$th element. According to the multiplication rule, its permutations are:

$$\underbrace{n \times n \times \cdots \times n}_{r} = n^r$$

### 19.5.1 Practice 8

1. How many ways are there to shoot 3 balls into 6 baskets?

2. If the all the digits can be used more than once, how many 3 digit numbers can be formed using the digits 0, 1, 2, 3, 4?

### 19.5.2 Exercise 19.5

1. The office documents of a company are coded with a combination of two of the first four letters of the alphabet $A, B, C, D$. If each letter can be used more than once, how many different codes can be formed?

2. There are eight paths from A to B. How many ways are there to go from A to B and then back to A?

3. A four digit number is to be formed with 4 numbers 3, 4, 5, 7, how many ways can it be formed such that its thousand digit and hundred digit are the same?

4. A basketball match has two possible result, either team $A$ wins or team $B$ wins. How many possible results are there if there are 10 matches?

5. A restaurant has introduced 7 sets of special dishes for lunch. There are 3 people come for lunch, each person orders one dish. How many different combinations of dishes can be ordered?

6. A building is assigned one security guard every night. If the security company has three assignable guards, how many different ways of assignments are there for a week?

7. A football match has three possible results, either team $A$ wins, team $B$ wins or the match is a draw. How many possible results are there if there are 8 matches?

8. A school is going to hold a trilingual speech competition. The rule stated that each class can only send one representative for each language. If there are 40 students in a class, how many different combinations of students can be chosen for the competition?

9. How many 6 digits odd number can be formed using the digitis 0, 1, 2, ⋯, 5 if each digit can be used more than once?

10. When there are only 7 digits, there are 8 ways to choose the first digit (2 - 10), 10 ways to choose the rest of the digits (0 - 10). Hence, there are $8\times10^6$ phone numbers. When there are 8 digits, there are 8 ways to choose the first digit, 10 ways to choose the rest of the digits. Hence, there are $8\times10^7$ phone numbers. Hence, there are:

    When there are only 7 digits, there are 8 ways to choose the first digit (2 - 10), 10 ways to choose the rest of the digits (0 - 10). Hence, there are $8\times10^6$ phone numbers. When there are 8 digits, there are 8 ways

11. The phone number format of a country is going from 7 digits to 8 digits. but the leading digit cannot be 0 or 1. How many new phone numbers will be formed after the change?

    **Sol.**

    When there are only 7 digits, there are 8 ways to choose the first digit (2 - 10), 10 ways to choose the rest of the digits (0 - 10). Hence, there are $8\times10^6$ phone numbers. When there are 8 digits, there are 8 ways to choose the first digit, 10 ways to choose the rest of the digits. Hence, there are $8\times10^7$ phone numbers. Hence, there are:

    $$\begin{aligned} 8\times10^7 - 8\times10^6 &= 80,000,000 - 8,000,000 \\ &= 72,000,000 \end{aligned}$$

    new phone numbers.

## 19.6 Combinations and Combination Formula

Grouping $r$ elements from $n$ elements ($r \le n$) without considering the order of the elements is called *combination* of $r$ elements from $n$ elements. The number of combinations of $r$ elements from $n$ elements is denoted by ${}^nC_r$, ${}_nC_r$, or $\binom{n}{r}$.

Take for example: taking two elements from three distinct element $a$, $b$, $c$ to form a group. Below are the combinations of any two elements from $a$, $b$, $c$:

| Permutations | $ab$ | $ba$ | $ac$ | $ca$ | $bc$ | $cb$ |
|---|---|---|---|---|---|---|
| Combinations | $ab$ | | $ac$ | | $bc$ | |

It can be considered as two steps:

First, combine any two elements from the three distinct elements to form a group. There are ${}_3C_2$ possible combinations.

Second, make full permutations for each combination. There are 2! possible permutations for each combination.

According to the multiplication rule,

$$\begin{aligned} {}_3P_2 &= {}_3C_2 \times 2! \\ \therefore\ {}_3C_2 &= \frac{{}_3P_2}{2!} \\ &= 3 \end{aligned}$$

Generalizing the above example, there are two steps to find the combinations of $r$ elements from $n$ elements:

First, combine any $r$ elements from the $n$ elements to form a group. There are ${}_nC_r$ possible combinations.

Second, make full permutations for each combination. There are $r!$ possible permutations for each combination.

According to the multiplication rule,

$$\begin{aligned} {}_nP_r &= {}_nC_r \times r! \\ \therefore\ {}_nC_r &= \frac{{}_nP_r}{r!} \\ &= \frac{n!}{(n-r)!r!},\ r \le n \end{aligned}$$

According to the definition ${}_nC_r = \frac{n!}{(n-r)!r!}$,

$$\begin{aligned} {}_nC_{n-r} &= \frac{n!}{(n-(n-r))!(n-r)!} \\ &= \frac{n!}{r!(n-r)!} \\ &= {}_nC_r \end{aligned}$$

That is,

$${}_nC_r = {}_nC_{n-r}$$

Note that:

1. If $r = n$, then ${}_nC_n = \frac{n!}{0!n!} = 1$.
2. If $r = 0$, then ${}_nC_0 = \frac{n!}{n!0!} = 1$.

### 19.6.1 Practice 9

1. There are six main cities in a country, each city has roads connecting to the other five cities. How many roads are there connecting the six cities?
2. There are 5 people in 4 cars, each car must have at least one person. How many ways are there to distribute the 5 people into 4 cars?

### 19.6.2 Practice 10

There are 4 different books. With the following criteria, how many ways are there to distribute the books?

1. Distribute evenly to two people.
2. Seperate evenly into two piles.

### 19.6.3 Exercise 19.6

1. Find the value of $n$ and $r$ of the following expessions:
   (a) ${}_{16}C_{r+3} = {}_{16}C_{7-r}$
   (b) ${}_{30}C_r = {}_{30}C_{r+2}$
   (c) ${}_nC_8 = {}_nC_7$
   **Sol.**

$$\begin{aligned} {}_nC_8 &= {}_nC_7 \\ {}_nC_r &= {}_nC_{n-r} \\ r &= 8 \\ n - r &= 7 \\ n - 8 &= 7 \\ n &= 15 \end{aligned}$$

2. Assume that $3({}_nC_4) = 5({}_{n-1}C_5)$, find the value of ${}_nC_9$.
3. There are 17 teams participating in a football competition. If each team plays against every other team, how many matches are there?
4. How many diagonal lines can be drawn in a convex nonagon?
5. There are 6 students on duty in a class, 1 student is in charge of cleaning the whiteboard, one is in charge of cleaning the rubbish bin, 2 students are in charge of sweeping the floor and 2 students are in charge of arranging the desks. How many ways are there to distribute the 6 students into the 4 jobs?
6. 4 people are to be choosen from 5 couples, where each couple cannot be chosen together. How many ways are there to choose the 4 people?
7. Three signal flags are to be chosen from 6 flags, one of which is red, two of which are yellow and the rest are blue. How many ways are there to choose the 3 flags?
8. A delegation with 6 members is to be formed from 9 students who major in Mathematics and 4 students who major in Education. With the following criteria, how many ways are there to form the delegation?
   (a) There are exactly two students who major in Education.
   (b) There are at least two students who major in Education.
9. Seperate 14 students evenly into 2 groups, how many ways are there to do so? How many ways are there to seperate these 14 students evenly into two classrooms?

## 19.7 Revision Exercise 19

1. Evaluate the following:

(a) $\frac{7!}{3!4!} + \frac{7!}{2!5!}$

(b) $\frac{{}^{11}P_5 + {}^{11}P_4}{{}^{12}P_5 + {}^{12}P_4}$

2. If $\frac{(n+6)!}{(n+4)!} = 18(n+1)$, find the value of $n$.

3. Assume that ${}_{2n}C_3 : {}_nC_2 = 44 : 3$, find the value of $n$.

4. All the letters from the word *TRIANGLE* are to be arranged. How many ways of arrangements are there such that the vowels are all seperated from each other?

5. One vowel and one consonant are to be chosen from the word *TRIANGLE* to form a word. How many ways are there to do so?

6. A basketball coach wants to choose two vanguards from four shooters, one center from two tall players, and two defender from four ball handlers. How many ways are there to form a team?

7. A committee must be formed by one lawyer, two engineers and two doctors chosen from 3 lawyers, 6 engineers and 7 doctors. How many ways are there to form the committee?

8. There is one of each note of the following value: \$1, \$2, \$5, \$10, \$20, \$50, \$100. How many distinct values of currency can be made from these notes?

9. There are 4 different history books, 5 different geography books and 3 different literature books on the shelf. If they are to be arranged in a row, how many ways are there to arrange them such that books of the same subject are arranged together?

10. 6 characters from 10 different characters $A$, $B$, $C$, $D$, $E$, and 0, 1, 2, 3, 4 are to be chosen to form a password. How many password can be formed such that there are no repeated characters and the password does not start with 0?

11. How many 7 digit even numbers can be formed from full permutations of the digits 0, 1, 2, 3, 4, 5, 6? How many multiple of 10 are there in these numbers?

12. How many permutations of all the characters in the word "fei li wu shi, fei li wu ting" are there?

13. How many ways are there to arrange all the letters in the word *ARRANGEMENT*?

14. Arranging 5 cups with different color and 5 canned juices with different flavour into a circle, how many ways are there to do so such that all the canned juices are not next to each other?

15. In a box of Chinese chess, there are 1 general, 2 advisors, 2 elephants, 2 horses, 2 chariots, 2 cannons and 5 soldiers of red color. Now we arrange these 16 pieces into a circle:

    (a) All the pieces of the same type are arranged together.

    (b) All the pieces of the same type are symmetrically arranged on one diameter.

16. If there is no repeated digits, how many odd numbers in between 4000 and 9000 can be formed from the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9?

17. How many 4 digit numbers with 2 repeated digits are there if the digits are chosen from 1, 2, 3, 4, 5 while the digits can be chosen more than once?

18. From the class committee of 10 members, at least 2 members, at most 8 members are to be chosen as representatives to attend a forum. How many ways are there to choose the representatives?

19. A team of at least 2 engineers and 1 technician is to be formed from 5 engineers and 4 technicians. How many ways are there to form the team?

20. 9 different books are to be distributed to 3 people $A$, $B$ and $C$. With the following conditions, how many ways are there to distribute the books?

    (a) Each people get 3 books.

    (b) $A$ gets 2 books, $B$ gets 3 books, $C$ gets 4 books.

    (c) One people get 2 books, one people get 3 books and one people get 4 books.

    (d) Seperated into 3 groups, each group has 3 books.

# Chapter 20

# Bionomial Theorem

## 20.1 Bionomial Theorem when $n$ is a Natural Number

Back in junior high school, we have learnt

$$(a+b)^1 = a+b$$
$$(a+b)^2 = a^2+2ab+b^2$$
$$(a+b)^3 = a^3+3a^2b+3ab^2+b^3$$

Now, let's discuss the expansion of $(a+b)^4$.

$$(a+b)^4 = (a+b)(a+b)(a+b)(a+b)$$

Each term in the expasion of $(a+b)^4$ is the product of one letter taken from each of the four brackets, that is to say, the expansion of should contain the following terms: $a^4$, $a^3b$, $a^2b^2$, $ab^3$, $b^4$.

Using the knowledge of permutation and combination, we can get the coefficient of each term in the expansion:

In the 4 brackets, if no $b$ is chosen, then there are ${}^4P_0$ ways to do so, therefore the coefficient of $a^4$ is ${}^4P_0$.

In the 4 brackets, if 1 $b$ is chosen, then there are ${}^4P_1$ ways to do so, therefore the coefficient of $a^3b$ is ${}^4P_1$.

In the 4 brackets, if 2 $b$ is chosen, then there are ${}^4P_2$ ways to do so, therefore the coefficient of $a^2b^2$ is ${}^4P_2$.

In the 4 brackets, if 3 $b$ is chosen, then there are ${}^4P_3$ ways to do so, therefore the coefficient of $ab^3$ is ${}^4P_3$.

In the 4 brackets, if all 4 $b$ is chosen, then there are ${}^4P_4$ ways to do so, therefore the coefficient of $b^4$ is ${}^4P_4$.

Therefore, $(a+b)^4 = {}_4C_0a^4 + {}_4C_1a^3b + {}_4C_2a^2b^2 + {}_4C_3ab^3 + {}_4C_4b^4$.

Generalize the above expansion, we have the following formula:

$$(a+b)^n = {}_nC_0a^n + {}_nC_1a^{n-1}b + \cdots + {}_nC_{n-r}ab^{n-r} + \cdots + {}_nC_nb^n$$

where $n \in \mathbb{N}$.

This formula is called the *Bionomial Theorem*, the polynomial in the right hand side is called the *Bionomial Expansion* of $(a+b)^n$, of which ${}_nC_0$, ${}_nC_1$, $\cdots$, ${}_nC_n$ are called the *Bionomial Coefficients*.

By looking at the formula above, we can know that,

1. The sum of the indices of $a$ and $b$ in each term is equal to the binomial expression, the index of $a$ decreases by 1 from $n$ to 0, while the index of $b$ increases by 1 from 0 to $n$.
2. The bionomial expansion has $n+1$ terms, that is to say, it has one term more than the exponent of the binomial expression.
3. Since ${}_nC_r = {}_nC_{n-r}$, therefore ${}_nC_0 = {}_nC_n$, ${}_nC_1 = {}_nC_{n-1}$, $\cdots$, ${}_nC_2 = {}_nC_{n-2}$, $\cdots$.

Bionomial expression can also be calculated by the following table:

| | | | | | | |
|---|---|---|---|---|---|---|
| $(a+b)^0$ | | | 1 | | | |
| $(a+b)^1$ | | 1 | | 1 | | |
| $(a+b)^2$ | 1 | | 2 | | 1 | |
| $(a+b)^3$ | 1 | 3 | | 3 | | 1 |
| $(a+b)^4$ | 1 | 4 | 6 | | 4 | 1 |
| $(a+b)^5$ | 1 | 5 | 10 | 10 | 5 | 1 |

$\vdots$

In the table above, for each row, except the beginning and the end being 1, all numbers except 1 are the sum of the two numbers above it, aka ${}_nC_{n+1}r = {}_nC_nr - 1 + {}_nC_nr$.

In bionomial theorem, let $a = 1$, $b = x$, then we get the following formula:

$$(1+x)^n = 1 + {}_nC_n1x + {}_nC_n2x^2 + \cdots + {}_nC_nrx^r + \cdots + x^n$$

### 20.1.1 Practice 1

Expand the following expression:

1. $(1+x)^7$
2. $(2+3x)^5$

### 20.1.2 Exercise 20.1

Expand the following expression (1 to 9):

1. $(m+n)^7$
2. $(3+2x)^4$
3. $(x-3)^5$
4. $(x+y^2)^6$

5. $\left(2+\frac{1}{x}\right)^5$
6. $\left(\frac{x}{3}+\frac{2}{x}\right)^4$
7. $\left(x-\sqrt[3]{x^2}\right)^3$
8. $\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)^6$
9. $(1+x+x^2)^3$
10. Calculate $\left(1+\sqrt{x}\right)^5+\left(1-\sqrt{x}\right)^5$.

## 20.2 General Form of Bionomial Expansion

In the bionomial expansion,

$$(a+b)^n = {}_nC_0a^n + {}_nC_1a^{n-1}b + \cdots + {}_nC_{n-r}ab^{n-r} + \cdots + {}_nC_nb^n$$

The $(r+1)$th term is

$$T_{r+1} = {}_nC_ra^{n-r}b^r$$

This is the general form of bionomial expansion.

### 20.2.1 Practice 2

Find the fourth term of $(x^3+2x)^7$ after expanding it in descending power of $x$.

### 20.2.2 Exercise 20.2

1. Find the coefficient of the fourth term of $(x+1)^9$ after expanding it in descending power of $x$.
2. Find the third term of $(3x+2)^5$ after expanding it in ascending power of $x$.
3. Find the middle term of $\left(1+\frac{x^2}{2}\right)^{10}$ after expanding it in ascending order of $x$.
4. Find the coefficient of $x^2$ in $(2-3x)^7$.
5. Find the constant term of $\left(x+\frac{1}{x}\right)^{10}$.
6. Find the coefficient of $\frac{1}{x^5}$ in the expansion of $\left(x-\frac{1}{x}\right)^9$.
7. Find the coefficient of $x^4$ in the expansion of $\left(2x+\frac{1}{\sqrt[3]{x}}\right)^8$.

## 20.3 Revision Exercise 20

1. Find the expansion of $(1-2x)^5$.
2. Expend $\left(2\sqrt{x}-\frac{1}{x}\right)^6$.
3. Find the eighth term of the bionomial expression $\left(\frac{3x^2}{2}-\frac{1}{3x}\right)^{11}$.
4. Find the middle term of $\left(x+\frac{1}{2\sqrt{x}}\right)^8$ after expanding it in descending power of $x$.
5. Find the coefficient of $x^{-12}$ in the expansion of the bionomial expression $\left(x^3-\frac{1}{x}\right)^{24}$.
6. If the coefficient of $x^4$ in the expansion of of the bionomial expression $(1+ax)^5$ is 80, find the value of $a$.
7. Given that the coefficient of the second, third, and fourth term if the expansion of $(1+x)^n$ after expanding it in ascending power of $x$ form an arithmetic progression, find the value of $n$.
8. Find the fourth term of $\left(px+\frac{q}{x}\right)^n$ after expanding it in descending power of $x$. If this is a constant term, find the value of $n$.

# Chapter 21

# Probability

In our daily life, a lot of stuff will yield certain results in certain conditions or situations. For example, by throwing a stone into the sky, the stone will fall down to the ground; the pure water will boil at 100°C. However, in some cases, there may be more than one possible result in a certain situation. For example, when we throw a coin into the air, it may land on the head or the tail, and the result is unpredictable. Nonetheless, if we do the experiment many times under the same conditions, we'll find certain patterns in the result after some analysis.

In order to find the pattern of a coin landing on the head, there are a lot of people who've conducted thousands of coin-tossing experiments, and here are the results:

| Experimenter | Tosses ($n$) | Heads ($m$) | Freq. $\left(\frac{m}{n}\right)$ |
|---|---|---|---|
| De Morgan | 2048 | 1061 | 0.5181 |
| Buffon | 4048 | 2048 | 0.5059 |
| Feller | 10000 | 4979 | 0.4979 |
| Pearson | 12000 | 6019 | 0.5016 |
| Pearson | 24000 | 12012 | 0.5005 |

From the results, we can see that when the number of tosses is large enough, the frequency $\frac{m}{n}$ of the coin landing on the head ($m$) will always be close to 0.5.

From that, we can see two obvious facts about this experiment:

1. Contingency: The result cannot be predicted in advance.
2. Inevitability: The results of the same experiment being conducted numerous times show a statistical regularity.

Probability theory is a branch of mathematics that studies statistical regularity in a mathematical way. In this chapter, we'll study the basic concepts of probability theory.

## 21.1 Sample Space and Events

Every possible results of a trial is called a *sample point* of the trial, and the set of all possible results is called the *sample space* of the trial, typically denoted by $S$. Take coin-tossing as an example, there are two possible resuls: head and tail. If we denote head by $H$ and tail by $T$, then the sample space of the coin-tossing experiment is $S = \{H, T\}$.

Although there are only two sample points in the coin-tossing experiment, there may be infinite sample points in some trials. For example, choose a number between 0 and 1, there will be an infinite amount of sample points, e.g. 0.1, 0.12, 0.145, etc.

### 21.1.1 Practice 1

1. Write down the sample space of throwing two coins once.
2. Write down the sample space of rolling a die once.
3. Select any number from 0, 1, ⋯, 9, and write down its sample space.
4. Write down the sample space of throwing a coin three times.

Within a trial, the set of a few sample points, that is, a subset of the sample space $S$, is called an *event* of the trial, and is usually denoted by capital letters $A$, $B$, $C$, etc. The sample space $S$ in itself is also an event that will surely happen, and is called the *sure event*. Empty set $\emptyset$ is also an event that will never happen, and is called the *impossible event* or *null event*. When an event only contains one sample point, that is, there is only one element of $S$ in the event, it is called a *simple event*.

For example, in the trial of throwing a dice for the dice points, the sure event $S = \{1, 2, 3, 4, 5, 6\}$, showing a dice points of 7 is an impossible event, and its denoted as $\emptyset$. The event of showing any of the dice points from 1 to 6 is a simple event.

Take another example, draw a card from a deck of 52 poker cards, there are 52 possible results. Hence, the sample space of this trial is a set of 52 elements, and any event is a subset of the sample space. Below are some example of events in this sample space:

1. The card drawn is a number 11 (impossible event)
2. The card drawn is a red heart 3 (simple event)
3. The card drawn is a number 9 (event with 4 elements)

4. The card drawn is a black spade (event with 13 elements)

5. The card drawn is not a club 5 (event with 51 elements)

Since the events are expressed as sets, listed below are some set operations used to describe relationships between events:

Let $A$ and $B$ be two events, then:

1. $A \cup B$ means that at least one of the events $A$ and $B$ will happen.

2. $A \cap B$ means that both events $A$ and $B$ will happen.

3. $A'$ means that the event $A$ will not happen.

### 21.1.2 Practice 2

Express the following events in set notation (1 to 4):

1. Throwing a dice, event $A$ = "showing a prime number dice points".

2. Throwing three dices, event $B$ = "total dice points less than 6".

3. Tossing two coins once, event $K$ = "showing exactly one head", event $L$ = "showing at least one head", event $M$ = "showing at most one head".

4. Tossing a coin three times:

   (a) $D$ = "get at least two heads".

   (b) $E$ = "the number of heads is lesser than the number of tails".

### 21.1.3 Exercise 12.1

1. Choose any two letters from the letters $K$, $O$, $T$, $A$, find:

   (a) The sample space $S$.

   (b) Event $A$ = "one letter is a vovel, the other is a consonant".

   (c) Event $B$ = "both letters are vovels".'

   (d) Event $C$ = "at least one letter is a vovel".

2. Throwing two dies, find the event where the sum of the dice points is a multiple of 3.

3. Throwing three dices, find the event where the sum of the dice points is 15.

4. Choose any two letters from the letters $A$, $B$, $C$, $D$, $E$ to form a row, find:

   (a) The sample space $S$.

   (b) Event $M$ = "there are exactly one vowel".

5. Throwing three coins, find:

   (a) The sample space $S$.

   (b) Event $A$ = "all coins show heads".

   (c) Event $B$ = "two coins show heads, one coin shows tail".

   (d) Event $C$ = "at least two coin shows heads".

6. Throwing two dices, express the following events in set notation:

   (a) $A$ = "the dice points of two dices are equal".

   (b) $B$ = "The dice points of one dice is twice the dice points of another".

   (c) $C$ = "The sum of the dice points is a multiple of 5 or 6".

## 21.2 Definition of Probability

In this section, we'll cover two definitions of probability, the classical and the statistical definition.

### Classical Definition of Probability

Doing a trial numerous times in the same conditions, the frequency of an event will show a certain regularity. Let's discuss the following example of dice-tossing:

Tossing a dice multiple time and recording the numbers of time of getting 1 dice point and the total number of tosses, we get the following results:

| Tosses ($n$) | No. of 1 dice points ($m$) | Frequency $\left(\frac{m}{n}\right)$ |
|---|---|---|
| 1000 | 174 | 0.1740 |
| 2000 | 350 | 0.1750 |
| 3000 | 499 | 0.1663 |
| 4000 | 673 | 0.1683 |
| 5000 | 837 | 0.1674 |
| 6000 | 9999 | 0.1665 |

From the table, we can see that as the number of tosses increases, the frequency of getting 1 dice point keep approaching a constant value of $\frac{1}{6} = 0.1667$.

When performing a large amount of repeated trials, the frequency of an event $A$ $\left(\frac{m}{n}\right)$ always approaches a constant

value. This constant value is called the *probability* of event $A$, and denoted as $P(A)$. This is the statistical definition of probability.

Since the occurrence of an event will never exceed the total number of trials, its frequency will always be a number between 0 and 1, that is, $0 \leq \frac{m}{n} \leq 1$. Hence, according to the statistical definition of probability, for any event $A$, its probability $0 \leq P(A) \leq 1$.

A sure event $S$ will always happen in every trial, so $P(S) = 1$. For an impossible event $\emptyset$, no matter how many trials we do, its occurrence will always be 0, so $P(\emptyset) = 0$.

## Classical Definition of Probability

Assume that a trial satisfies the following conditions:

1. The outcome of the trial is finite.
2. The probability of each outcome is equal.

This kind of trial model is called a *classical probabilistic model*.

Let $S$ be the sample space of the trial that contains $n$ equally probable, $A$ is an event that contains $m$ outcomes, then the probability of event $A$ is:

$$P(A) = \frac{m}{n} = \frac{n(A)}{n(S)}$$

This is the classical definition of probability.

### 21.2.1 Practice 3

1. Randomly draw two cards from a deck of 52 poker cards, find the probability of two "$K$" cards.
2. Between all students in a class, there are 13 students with type A blood, 10 students with type B blood, 2 students with type AB blood, and 15 students with type O blood. If we randomly choose 4 students, find the probability of the following events:
   (a) 2 type A blood, 2 type B blood
   (b) 2 type A blood, 1 type AB bloog, 1 type O blood
   (c) 1 for each type of blood.
3. Casually arranging all the letters in the word *GERMANY*, find the probability of the following events:
   (a) 5 adjacent consonants
   (b) 5 non-adjacent consonants

### 21.2.2 Exerise 21.2

1. A bag contains 9 balls, of which 2 are white, 3 are red, and 4 are yellow. Randomly drawing one ball, find the probability of the following events:
   (a) The ball drawn is red
   (b) The ball drawn is not red
   (c) The ball drawn is yellow
2. A box contains 3 throat lozenges and 5 bubble gums. Randomly drawing two of them, find the probability of getting two bubble gums.
3. Randomly drawing 3 cards from a deck of 52 poker cards, find the probability of getting 3 cards of spades.
4. There are 4 novels and 8 essay collections in a box. Randomly drawing 3 books from the box, find the probability of the following events:
   (a) All three books are novels
   (b) Two books of novels, one book of essay collection
   (c) All three books are essay collections
5. A committee is to be formed by selecting 4 people from 6 males and 5 females. Find the probability of the following events:
   (a) All 4 people are male
   (b) All 4 people are female
   (c) 2 people for each gender
6. Of 100 products, 95 are quality products, and 5 are defective products. Randomly drawing 2 products, find the probability of the following events:
   (a) All 2 products are quality products
   (b) All 2 products are defective products
   (c) 1 quality products, 1 defective product
7. Randomly choose 3 letters from the word *TRIANGLE*, find the probability of the following events:
   (a) More vowels than consonants
   (b) More consonants than vowels
8. Tossing three dices at the same time once, find the probability of the sum of dice points larger than 15.
9. Randomly shuffling each digits in 2233344455, find the probability of two 2s being adjacent to each other.

10. Randomly shuffling 5 cards that are assigned to 1, 2, 3, 4, 5 respectively and putting them into a 5 digit number, find the probability of the number being an even number.

## 21.3 Addition Rule

### Mutually Exclusive Events and Inclusive Events

When two events $A$ and $B$ can happen at the same time, event $A$ and $B$ are said to be *inclusive*. For example, when tossing a dice once, the events "getting an even number" and "getting a multiple of 3" can happen at the same time. Therefore, these two events are inclusive events. Now let's find the probability of getting an even number or a multiple of 3 when tossing a dice once.

Let the sample space of the trial be $S = \{1, 2, 3, 4, 5, 6\}$, $n(S) = 6$.

Let event $A$ = "getting an even number" $= \{2, 4, 6\}$, $n(A) = 3$, therefore $P(A) = \frac{3}{6}$.

Let event $B$ = "getting a multiple of 3" $= \{3, 6\}$, $n(B) = 2$, therefore $P(B) = \frac{2}{6}$.

$A \cup B$ = "getting an even number or a multiple of 3" $= \{2, 3, 4, 6\}$, $n(A \cup B) = 4$, therefore $P(A \cup B) = \frac{4}{6} = \frac{2}{3}$.

$A \cap B$ = "getting an even number that is also a multiple of 3" $= \{6\}$, $n(A \cap B) = 1$, therefore $P(A \cap B) = \frac{1}{6}$.

Generally speaking, according to the formula of cardinality of the union of two sets,

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

Dividing both side by $n(S)$, we get:

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)}$$

That is,

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

The relationship is the *addition rule* of probability.

In the example above,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{3}{6} + \frac{2}{6} - \frac{1}{6} \\ &= \frac{2}{3} \end{aligned}$$

When two events $A$ and $B$ cannot happen at the same time, that is, $A \cap B = \emptyset$, $A$ and $B$ is said to be mutually exclusive. For example, there are a red ball and a white ball in a bag. When randomly drawing one ball, the ball will be either red or white, but not both. This is a mutually exclusive event.

Since $A$ and $B$ are mutually exclusive, $P(A \cap B) = 0$.

$$P(A \cup B) = P(A) + P(B)$$

The relationship above is the additional rule of mutually exclusive events.

From that, we know that if event $A$ and event $B$ are mutually exclusive, their probability is the sum of their individual probabilities.

### 21.3.1 Practice 4

1. A bag contains 5 cards each for the color red, blue and green. Randomly drawing one card, find the probability of getting a red card or a yellow card.

2. Drawing a card from a deck of 52 poker cards, find the probability of getting a heart or number that is a multiple of 5.

3. Of a class of 45 students, 30 of them has visited Bali or Jakarta, of which 14 of them has visited Bali and 10 of them has visited Jakarta. Randomly choosing one student, find the probability of the student has visited both cities.

### Complementary Events

During a trial, if one of two mutually exclusive events $A$ and $B$ must happen, then these two events are said to be complementary events. The complementary event of an event $A$ is denoted by $A'$. Take coin-tossing as an example, if event $A$ is "getting a head", then its complementary event $A'$ is "getting a tail". Since either event $A$ or $A'$ must happen, therefore $A \cup A' = S$. Applying additional rule for mutually exclusive events,

$$\begin{aligned} P(A) + P(A') &= P(A \cup A') \\ &= P(S) \\ &= 1 \\ \therefore P(A) &= 1 - P(A') \end{aligned}$$

The relationship above can be used to calculat the probability of complementary events.

### 21.3.2 Practice 5

1. There are 22 boys and 23 girls in a class. Randomly picking 2 students from them, find the probability of having at least one boy in these 2 people.

2. In a lucky draw, there are 60 boxes, of which 5 of them contain a prize. Randomly picking 2 boxes, find the probability of getting at least one prize.

### 21.3.3 Exercise 21.3

1. There are 5 red balls, 6 yellow balls, and 8 black balls in a box. Randomly drawing one ball, find the probability of getting a red ball or a yellow ball.

2. There are 18 different reference books on the shelf, of which 6 of them are Chinese and 5 of them are Maths, and the rest of them are Ecomoics. Randomly picking two books from the shelf, find the probability of getting at least one Math book or one Chinese book.

3. There are 15 shirts, 10 vests and 5 T-shirts on the rack. Randomly picking two clothes from the rack, find the probability of getting at least one T-shirt.

4. There are 50 prizes in a lucky draw, of which 1 of them is cash prize worth \$800, 2 of them are cash prize worth \$500, 5 of them are cash prize worth \$100, and the rest of them are bookshop vouchers worth \$10. One person make two draws, find the probability of getting at least one cash prize.

5. There are 50 people doing a meeting in a classroom, of which 35 of them are students, 12 of them are parents, and 3 of them are teachers. Randomly picking one spokesperson, find the probability of the following events:

   (a) The spokesperson is either a teacher or a student.

   (b) The spokesperson is either a teacher or a parent.

   (c) The spokesperson is either a student or a parent.

6. There are 100 lottery tickets, of which 3 of them are winning tickets. One person has bought 10 tickets, find the probability of the following events:

   (a) All tickets are not winning tickets.

   (b) At least one ticket is a winning ticket.

7. There are 11 out of 45 students in a class who have donated their blood before. Randomly picking three students, find the probability of the following events:

   (a) All three students have donated their blood before.

   (b) All three students have never donated their blood before.

   (c) At least one student has donated his/her blood before.

8. Tossing two dices at the same time, find the probability of the sum of the dice points being 9.

9. Continuously tossing a coin 5 times, find the probability of getting at least one head.

10. There are 7 red balls and 10 white balls in a box. Randomly drawing 3 balls, find the probability of the following events:

    (a) Getting at least one red ball.

    (b) Getting at least one white ball.

    (c) Getting at least two red balls.

11. Tossing 3 dices once, find the probabily of hte following events:

    (a) Exactly one dice shows 6.

    (b) Exactly one dice or all three dices show 1.

## 21.4 Multiplication Rule

### Independent Events

For two events $A$ and $B$, if the occurrence of event $A$ does not affect the occurrence of event $B$, then the two events are said to be independent events. For example, tossing a dice twice, the result of the first toss does not affect the result of the second toss, vice versa. Therefore, the two events are independent events. The probability of the two events occurring is the product of their individual probabilities:

$$P(A \cap B) = P(A) \times P(B)$$

### 21.4.1 Practice 6

1. There are 10 oranges and 12 apples in two baskets respectively, of which 2 oranges and 4 apples are rotten. Randomly picking one fruit from each basket, find the probability of getting two rotten fruits.

   **Sol.**

   The probability of getting a rotten orange is $\frac{2}{10} = \frac{1}{5}$.
   The probability of getting a rotten apple is $\frac{4}{12} = \frac{1}{3}$.

Therefore, the probability of getting two rotten fruits is $\frac{1}{5} \times \frac{1}{3} = \frac{1}{15}$.

2. Person $A$ and person $B$ are participating in a archery competition. Person $A$ and person $B$ have a probability of $\frac{5}{8}$ and $\frac{7}{15}$ respectively to hit the bullseye. Find the probability the following events:

   (a) Both person $A$ and person $B$ hit the bullseye.

   **Sol.**

   Let the event of person $A$ and person $B$ both hitting the bullseye be $A$.

   $$\begin{aligned} P(A) &= \frac{5}{8} \times \frac{7}{15} \\ &= \frac{7}{24} \end{aligned}$$

   (b) Both of them does not hit the bullseye.

   **Sol.**

   Let the event of person $A$ not hitting the bullseye be $B$, then $B'$ is the event of person $A$ hitting the bullseye.

   Let the event of person $B$ not hitting the bullseye be $C$, then $C'$ is the event of person $B$ hitting the bullseye.

   $$\begin{aligned} P(B \cap C) &= P(B) \times P(C) \\ &= 1 - P(B') \times 1 - P(C') \\ &= 1 - \frac{5}{8} \times 1 - \frac{7}{15} \\ &= \frac{3}{8} \times \frac{8}{15} \\ &= \frac{1}{5} \end{aligned}$$

   (c) At least one of them hits the bullseye.

   **Sol.**

   Let the event of at least one of them hits the bullseye be $D$, then $D'$ is the event of both of them not hitting the bullseye.

   $$\begin{aligned} P(D) &= 1 - P(D') \\ &= 1 - \frac{1}{5} \\ &= \frac{4}{5} \end{aligned}$$

### 21.4.2 Exercise 21.4a

1. Tossing a dice three times, find the probability of getting 2 for the first and the second toss, and an odd number for the third toss.

2. Tossing a coin 5 times, find the probability of getting at 4 heads continuously.

3. The forecast accuracy of a weather station is 80%. Find probability of getting 4 accurate forecasts our of 5 forecasts. (Round to the nearest 4 decimal places)

4. There are three processes for processing a component, the failure rate of each process are 1.2%, 1.8%, and 0.8% respectively. If the chance of failure of the component does not depend on other processes, find the success rate of the component.

5. Randomly drawing 1 card from a deck of 52 cards, and draw another card without putting the first card back into the deck. Find the probability of getting a spade for the second card.

6. 3 people are decrypting a message on their own. The probability of each of them decrypting the message correctly is $\frac{1}{5}$, $\frac{1}{4}$ and $\frac{1}{3}$ respectively. Find the probability the message getting decrypted.

7. 3 people are participating in an archery competition. person $A$ make three hits out of five, person $B$ make two hits out of three, person $C$ make one hit out of two. Now each person gets one shot, find the probability of the following events:

   (a) All three of them make a hit.

   (b) Only one of them make a hit.

   (c) At least one of them make a hit.

8. There are 5 black balls, 4 yellow balls, and 3 white balls in a box. Randomly draw one balls and put it back into the box. Repeat the process 3 times. Find the probability of the following events:

   (a) Only get one yellow ball

   (b) All three balls are yellow balls

   (c) Get one black ball, one yellow ball, and one white ball

### 21.4.3 Practice 7

There are 80% of the families living in a town have signed up for fibre optic cable. Now choose 10 random families from the town for a survey, find the probability of the following events (round to the nearest 4 decimal places):

1. Exactly half of the families have signed up for fibre optic cable.

2. All of the families have signed up for fibre optic cable.

3. At least one family has not signed up for fibre optic cable.

### 21.4.4 Exercise 21.4b

1. The survival rate of a plant is 0.6. Now 10 plants are planted, find the probability exactly 5 of them survive.

2. The hit rate of a person shooting a basket ball is 0.4. Find the probability of him making 10 hit out of 25 shots.

3. According to statistical data, there are 85% of the population in a city has Hepatitis B Vaccination. Now randomly choose 8 people from the city for a health check, find the probability of at most two of them have not been vaccinated.

4. The probability of a medicine successfully curing a cold is 0.96. Now 10 cold patients are taking the medicine, find the probability of at least 8 of them are cured.

5. A factory produces a component, the probability of the component being defective is 0.04. Now 20 components are produced, find the probability of the following events:

   (a) Exactly 1 of them is defective.

   (b) Exactly 2 of them are defective.

   (c) At most one of them is defective.

6. Tossing a coin 5 times, find the probability of the following events:

   (a) Exactly 3 heads.

   (b) At least 3 heads.

   (c) The number of heads is odd.

7. A person will4 four traffic lights on his way to work. Given that the duration of red, yellow, and green lights are $90s$, $5s$ and $25s$ respectively for each light. Find the probability of the following events:

   (a) Get a red light for the every light.

   (b) Only get a red light for the first two lights.

   (c) Get exactly two red lights.

## 21.5 Mathematical Expectation

Consider the following scenario: during a commercial activity, the probability of a person getting a profit of \$300 is 0.6, and the probability of a person getting a loss of \$100 is 0.4. He did the activity 10 times. According to the probability, this guy was expected to get a profit of \$300 for 6 times, and a loss of \$100 for 4 times. Therefore, the expected profit of these 10 times of commercial activities is $300\times6+(-100)\times4$.

$$\begin{aligned}\text{Average Profit} &= \frac{300\times6+(-100)\times4}{10}\\ &= \frac{1800-400}{10}\\ &= \frac{1400}{10}\\ &= 140\end{aligned}$$

This average value is called the *mathematical expectated value*, or *expected value* for short, of this people doing the commercial activity.

The expect value does not mean that the actual profit of the person will be \$140 for every activity he did. It is just the average value of the profit he will get if he did the activity a large number of times.

Generalize the above example, let the probability of someone getting a profit of $\$x_1$, $\$x_2$, $\cdots$, $\$x_k$ be $p_1$, $p_2$, $\cdots$, $p_k$ respectively, of which $p_1+p_2+\cdots+p_k=1$. Then the expected value of the profit is

$$E = x_1p_1 + x_2p_2 + \cdots + x_kp_k$$

### 21.5.1 Practice 8

1. One person pays \$2 to play a a game. The chance of winning the game is 0.4, and he will get \$3 if he wins. Find the expected value of profit of this person in this game.

   **Sol.**

   The chance of winning the game is 0.4, and the profit of winning the game is \$1;

   The chance of losing the game is 0.6, and the loss of losing the game is \$2.

   Therefore, the expected value of profit of this person in this game is

   $$\begin{aligned}E &= 0.4\times1+0.6\times(-2)\\ &= 0.4-1.2\\ &= -\$0.80\end{aligned}$$

2. During an investment activity, the probability of a person getting a profit of \$2,500 is 0.55, and the probability of a person getting a loss of \$1,200 is 0.45. Find the expected value of profit of this person in this investment activity.

   **Sol.**

   The expected value of profit of this person in this investment activity is

$$\begin{aligned} E &= 2500 \times 0.55 + (-1200) \times 0.45 \\ &= 1375 - 540 \\ &= \$835 \end{aligned}$$

### 21.5.2 Exercise 21.5

1. During a commercial activity, the probability of a person getting a profit of $\$10,000$ is $\frac{3}{5}$, and the probability of a person getting a loss of $\$6,000$ is $\frac{2}{5}$. Find the expected value of profit of this person in this commercial activity.

2. A company produces light bulbs. The profit of producing a quality light bulb is \$2, and the loss of producing a defective light bulb is \$0.50. Assume that the chance of producing a defective light bulb is 0.02. Find the expected value of profit of this company in producing each light bulb.

3. A company has insured accident insurance worth \$60 for its employees, and the employees will get $\$1,200$ if they are involved in an accident. Given that the probability of an an accident happening is 0.005, find the expected value of profit of the insurance company in this insurance.

4. An airline provides an \$8 aviation assurance plan to its passengers. If flight is delayed for more than an hour, the passengers will get \$250 compensation. Given that the percentage of a flight being delayed for more than an hours is 2%, find the expected value of profit of the airline.

5. The price of a lottery ticket is \$2. The probability of winning the lottery is as follows: $\frac{1}{10000}$ for $\$5,000$, $\frac{1}{1000}$ for \$500. Find the expected value of someone who buys a lottery ticket.

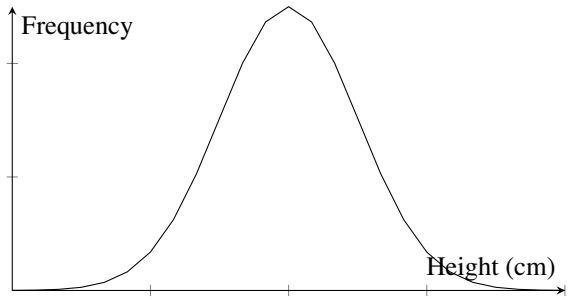
6. A person pays \$1 to play a game. The probability of him getting $\$3,000$, $\$2,000$, and $\$1,000$ are all $\frac{1}{10000}$. Find the expected value of this person in this game, and determine of this game is worth playing.

7. A high school has released 8000 \$1 lottery tickets, 5 of which has a prize of \$500, 8 of which has a prize of \$300, 10 of which has a prize of \$100, and 50 of which has a prize of \$10. Find the expected value of someone who buys a lottery ticket.

## 21.6 Normal Distribution

In the probability models that we have discussed in the previous sections, the number of results of a trial is limited, that is, the limited sets in the sample space. However, there are many situations in which the results are real numbers within a certain range.

For example, measuring the height of senior 2 boy students, the results (in $cm$) is a real number bigger than 0. As the number of students getting measured become larger and larger, the frequency polygon of the height of the students will become a bell curve as shown in the figure below. This bell curve is called the *normal curve*.



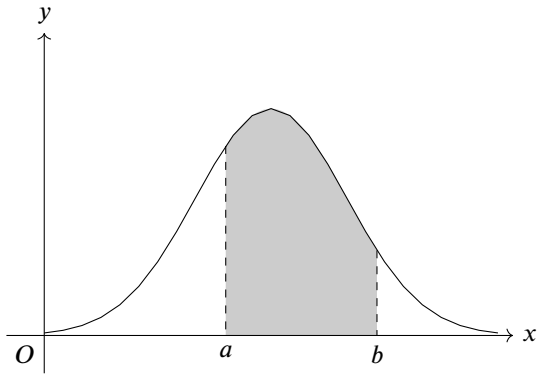

The normal distribution is the most common probability model in real life. A normal distribution consists of two parameters, the mean value $\mu$ and the standard deviation $\sigma$, thier corresponding functional expression of normal curve is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Assume that the observation $X$ of a trial is normal distribution with mean value $\mu$ and standard deviation $\sigma$, then its denoted as:

$$X \sim N(\mu, \sigma^2)$$

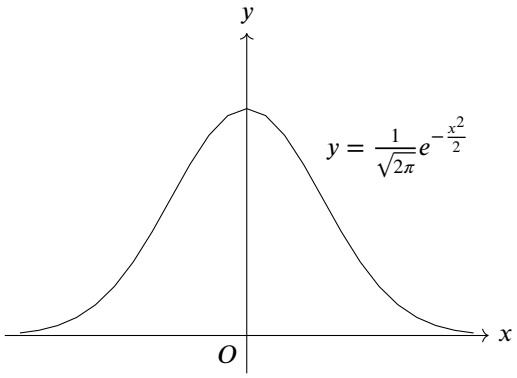
The probability $P\ (a < X < b)$ of the observation $X$ being in the interval $(a, b)$ is the area formed by the x-axis, line $x = a$, line $x = b$, and the normal curve, as shown in the figure below.




A normal distribution with mean value of 0 and standard deviation of 1 is called the standard normal distribution, its normal curve

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

is symmetrical about y-axis, as shown in the figure below.




Assume that $X$ is a normal distribution with mean value $\mu$ and standard deviation $\sigma$, then

$$Z = \frac{X - \mu}{\sigma}$$

is a standard normal distribution. That is,

$$\text{If } x \sim N(\mu, \sigma^2), \text{ then } Z = \frac{x - \mu}{\sigma},\ Z \sim N(0, 1)$$

Hence, General problems of normal distribution can be solved by converting into standard normal distribution problems.

In order to calculate the probability of events related to standard normal distribution, we have attached a standard normal distribution table in appendix A. Assume that $Z$ is a standard normal distribution, $z$ is a positive number lesser than or equal to 3.49, we can use the table to calculate the probability of $Q(z) = P(Z \leq z)$.

For example, to find $P(Z \leq 1.23)$, first find the row of $z = 1.2$. The value in the column of this row corresponding to 3 (2nd decimal place) is 0.1093, this is the probability of $P(Z \leq 1.23)$.

| $z$ | 0 | 1 | 2 | 3 | ⋯ |
|---|---|---|---|---|---|
| | | | | | |
| 1.2 | 0.1151 | 0.1131 | 0.1112 | **0.1093** | |
| | | | | | |

Any probability of events related to standard normal distribution can be solved using symmetry and properties of probability.

Since the probability of sure event is 1, the area of the region under the curve of standard normal distribution above the x-axis is 1.

Assume that $X$ is a normal distribution, $x$ is any real number, $P(X = x) = 0$. Therefore, $P(X < x) = P(X \leq x)$, $P(X > x) = P(X \geq x)$.

### 21.6.1 Practice 9

There are 2500 students who have attended geography exam during the Senior UEC exam. their marks can be assumed to be a normal distribution with mean value of 60 and standard deviation of 11.

(a) Find the number of students who failed the exam.

(b) If grade $A1$ is 78 marks or above, find the percentage of students who obtained grade $A1$.

(c) If the passing rate is 90%, find the minimum marks required to pass the exam. (Round to the nearest integer)

### 21.6.2 Exercise 21.6

1. If the observation of a trial $Z$ is a standard normal distribution, find the probability of the following events:
   (a) $P(Z < 0.91)$
   (b) $P(Z \leq -2.01)$
   (c) $P(Z \geq -0.5)$
   (d) $P(0.24 < Z \leq 1.79)$
   (e) $P(-2.21 < Z < -0.97)$
   (f) $P(-2.39 < Z \leq 0.56)$
2. If the observation of a trial $Z$ is a standard normal distribution, find the value of $a$ of the following events:
   (a) $P(Z > a) = 0.0505$
   (b) $P(Z < a) = 0.8980$
   (c) $P(Z < a) = 0.3632$
   (d) $P(Z > a) = 0.8599$
   (e) $P(|Z| > a) = 0.0142$
   (f) $P(|Z| < a) = 0.7888$

3. A factory produces canned coffee. The capacity of each can can be assumed to be a normal distribution with mean value of $244.5ml$ and standard deviation of $5.4ml$.

   (a) Randomly pick a can of coffee, find the probability that the capacity of the can is less than $235ml$.

   (b) Randomly pick a can of coffee, find the probability that the capacity of the can is in between $240ml$ and $250ml$.

   (c) The factory has sold $18,000$ can of coffee in a week. How many cans have a capacity greater than $260ml$?

   (d) Someone has bought 7 cans of coffee, find the probability that 4 cans of coffee have less than $2420ml$ of capacity each.

4. A machine produces wrapped cookies, the weight of each cookie can be assumed to be a normal distribution with mean value of $501.25g$ and standard deviation of $5.32g$.

   (a) Find the percentage of cookies that have a weight less than $490g$.

   (b) Of the $10,000$ cookies produced by the machine, how many cookies have a weight greater than $510g$?

   (c) If there are 2.5% of the cookies that have a weight less than $a$ grams, find the value of $a$.

   (d) If there are 99.02% of the cookies that have a weight between $(501.25-c)$ grams and $(501.25+c)$ grams, find the value of $c$.

5. The life span of brand $A$ television can be assumed to be a normal distribution with mean value of 7.28 years and standard deviation of 2.23 years. The warranty of the television is 3 years. If the television is broken within the warranty period, the company will replace it with a new one for free. Find the percentage of television that will be replaced.

6. The number of letters received by a company in a working day. Given that the probability of receiving more than 150 letters in a working day is 0.1210, while the probability of receiving more than 50 letters in a working day is 0.9495. Find the mean value and standard deviation of the number of letters received by the company in a working day.

7. There are $2,256$ students sitting on the entrance exam for a high school. THe full marks of the exam is 400 marks, and there are $1,200$ students who get accepted into the high school. The marks of the students can be assumed to be a normal distribution with mean value of 189 marks and standard deviation of 53.

   (a) Find the number of students who get less than 160 marks.

   (b) Find the lowest marks that a student can get to be accepted into the high school. (Round to the nearest integer)

   (c) If the students who get more than 300 marks are eligible to get a scholarship, find the number of students who are eligible to get a scholarship, and their percentage.

   (d) The school administration stipulates that the students who get less than 200 marks has to take a remedial class. Find the number of students who have to take a remedial class.

8. A group of data is normally distributed with mean value of $\mu$ and standard deviation of $\sigma$. Find in this group of data:

   (a) The percentage of data that is in between $\mu-\sigma$ and $\mu+\sigma$.

   (b) The percentage of data that is in between $\mu-2\sigma$ and $\mu+2\sigma$.

   (c) The percentage of data that is in between $\mu-3\sigma$ and $\mu+3\sigma$.

## 21.7 Revision Exercise 21

1. Tossing two dices at the same time, find the event where the sum of the two dices is 8.

2. There are 30 boys and 12 girls in a class. Randomly pick a representative, find the probability of a boy getting picked.

3. Randomly pick a card from a deck of 52 poker card, find the probability of the following events:

   (a) Getting a spades.

   (b) Not getting a spades.

4. Randomly pick a card from a deck of 52 poker card, find the probability of the following events:

   (a) Getting a club.

(b) Getting an ace.

(c) Getting a club ace.

(d) Getting a club or an ace.

5. Randomly write down a two digit number, find the probability of the following events:

   (a) Larger than 20.

   (b) An even number.

   (c) An odd number.

6. Let a phone number consist of 7 digits formed by 0, 1, 2, $\cdots$, 9. Someone only remembers the first three digits and forgots the last four digits of the phone number of his colleague. Find the probability him calling the right person with just one dial.

7. Person $A$ and $B$ toss two dices each at the same time. Person $B$ gets 10 points, what is the probability of person $A$ winning?

8. In a archery competition, the probability of person $A$ and person $B$ winning are 0.35 and 0.45 respectively. Find the probability of both of them losing.

9. There are 200 staff members in a company, a quarter of which are foreigners. There are 115 male staff members and 85 female staff members in the company. Given that 20 female staff members are foreigners. Now randomly pick a staff member, find the probability of the staff member picked is male and a native.

10. Drawing three cards from a deck of 52 poker card, find the probability of drawing at least one face card.

11. The hit rate of a person shooting a basketball is 0.8. If he shoots three times, find the probability of him scoring at least two times.

12. The accuracy of forecast of a weather station is 89%. Find the probability of five accurate forecasts in a week.

13. Given that the probability of a 18-year-old teenager being drawn for national service is 0.2. Given that a community has 4 18-year-old teenagers, find the probability of at least one of them being drawn for national service.

14. Tossing a dice, getting a number 6 can get \$30, while getting other numbers can get \$3. Find the expected value of the game.

15. There are 4 50 cent coins and and 6 20 cent coins in a bag. A person randomly pick two coins from the bag. Find the expected value he gets.

16. A food stall perpares 250 packets of nasi lemak every day. The cost of each packet is \$1.50, and the selling price is \$5.00, unsold packets are thrown away. According to statistical data, the stall can sell 57% of the nasi lemaks. Find the expected value of the profit of the stall.

17. In a lucky draw, there are 15 envelopes with cash prizes inside, one of which has \$100, two of which has \$50, three of which has \$10, four of which has \$5, and five of which has \$1. One person draw one envelope from the lucky draw, find the expected value. If the person pays \$15 to draw one envelope, determine whether it is worth it to pay for the lucky draw.

18. The winning rate of a guessing game in a charity fair are as follows: winning the probabilities of \$2,000, \$500, \$200 are all $\frac{1}{5000}$, while the probability of winning \$150 is $\frac{1}{3000}$. If the fee to play the game once is \$1, find the expected value of the game. Is it worth it for the player to play the game?

19. The weight distribution of 2,524 girls in a school can be assumed to be a normal distribution with mean weight of $53.79kg$ and standard deviation of $7.24kg$.

   (a) Randomly pick a girl, find the probability of her weight being less than $40kg$.

   (b) Find the number of girls whose weight is greater than $65kg$.

   (c) Find the percentage of girls whose weight is between $45kg$ and $55kg$.

   (d) Assume that there are 10.03% of the girls whose weight is greater than $c$kg, find $c$.

   (e) Randomly pick 10 girls, find the probability of at least two of them having weight lesser than $55kg$.

20. The duration of phone calls of customer service of a company can be assumed to be a normal distribution. Given that of all the phone calls, 1.02% of them are more than 30 minutes long, 25.14% of them are less than 20 minutes long. Find the mean value and standard deviation of the duration of phone calls.

# Appendix A

# Standard Normal Distribution Table

Listed in the table below are $Q(z) = P(Z \leq z)$, in which $Z$ is a standard normal distribution $N(0, 1)$.

| $z$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.4960 | 0.4920 | 0.4880 | 0.4840 | 0.4801 | 0.4761 | 0.4721 | 0.4681 | 0.4641 |
| 0.1 | 0.4602 | 0.4562 | 0.4522 | 0.4483 | 0.4443 | 0.4404 | 0.4364 | 0.4325 | 0.4286 | 0.4247 |
| 0.2 | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| 0.3 | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| 0.4 | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| 0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |
| 0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| 0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| 0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| 0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| 1.0 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| 1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| 1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| 1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| 1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| 1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| 1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| 1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |
| 1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| 1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| 2.0 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| 2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| 2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0125 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| 2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| 2.4 | 0.0082 | 0.0080 | 0.0078 | 0.0075 | 0.0073 | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| 2.5 | 0.0062 | 0.0060 | 0.0059 | 0.0057 | 0.0055 | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| 2.6 | 0.0047 | 0.0045 | 0.0044 | 0.0043 | 0.0041 | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| 2.7 | 0.0035 | 0.0034 | 0.0033 | 0.0032 | 0.0031 | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| 2.8 | 0.0026 | 0.0025 | 0.0024 | 0.0023 | 0.0023 | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| 2.9 | 0.0019 | 0.0018 | 0.0018 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| 3.0 | 0.0013 | 0.0013 | 0.0013 | 0.0012 | 0.0012 | 0.0011 | 0.0011 | 0.0011 | 0.0010 | 0.0010 |
| 3.1 | 0.0010 | 0.0009 | 0.0009 | 0.0009 | 0.0009 | 0.0008 | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| 3.2 | 0.0007 | 0.0007 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0005 |
| 3.3 | 0.0005 | 0.0005 | 0.0005 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0003 |
| 3.4 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0002 |