Mathematics in the Modern World

Lecture Notes

Rey Joseph Nieto

Version: 22 August 2026

Preface

I prepared these lecture notes for a class in *Mathematics in the Modern World* composed entirely of biology majors. The discussion, examples, and exercises are therefore written with that audience in mind. I want the notes to do more than help students complete a required mathematics course. The ideas and habits of reasoning developed here should remain useful throughout university and in later scientific, professional, and everyday work. Where practicable, the notes connect mathematics with biology, health, data, technology, and decisions that future biologists and healthcare professionals may have to make.

I have followed the curriculum guidelines for *Mathematics in the Modern World* issued by the Philippine Commission on Higher Education (CHED), including the prescribed learning outcomes and subject coverage [1]. Within those requirements, I have chosen explanations, applications, and supplementary examples for biology students. CHED has not reviewed or endorsed these independently prepared notes.

The general coverage of the course is now settled, but these notes remain a work in progress. Later versions may add examples, clarify explanations, or improve the exposition without changing the overall scope. Reports of errors and suggestions for corrections or improvements are welcome. Please send them to

`rj (at) rjnieto (dot) com`

Course Access and Assessment. I want the assessments in this course to measure mathematical understanding, not proficiency in a particular language or access to a device. Whenever practicable, you may explain your reasoning in English, Tagalog, Cebuano, or Taglish. If you do not have a suitable device, you may request printed course notes free of charge.

Clear instructions, appropriate access arrangements, and prompt correction of marking or recording errors matter for the same reason. A mark affected by unclear instructions or an unmet access need no longer measures mathematical understanding alone.

Other instructors, including those at other universities, are welcome to use, adapt, and redistribute these notes under the MIT License reproduced below. Please attribute the work to Rey Joseph Nieto and retain the copyright and permission notice in copies or substantial portions of the notes.

MIT License for These Lecture Notes

Contents

Chapter 1

Mathematics in Our World

Mathematical ability is often equated with numerical fluency. The association is understandable: much of school mathematics concerns arithmetic, algebra, trigonometry, statistics, probability, and calculus. Numerical skill matters, but mathematics extends beyond calculation. It studies patterns, structures, quantities, shapes, change, relationships, and uncertainty. It also supplies a language for describing these subjects and a logical framework for reasoning about them.

Learning Outcomes

After completing this chapter, you should be able to:

- identify examples of patterns and regularities in nature and everyday life;
- explain why mathematics is more than numerical calculation;
- distinguish a mathematical model from the reality it represents;
- recognise examples of inductive and deductive reasoning;
- describe some practical, intellectual, and aesthetic dimensions of mathematics; and
- explain how mathematics contributes to biology and modern technology.

Planning an ordinary morning. Some uses of mathematics in everyday life are fairly obvious. Suppose your first class begins at 8:00 a.m. What time should you wake up to arrive on time? You might estimate how long it takes to get out of bed, take a shower, get dressed, prepare your belongings, eat breakfast, and travel to school. At first, this appears to be a simple arithmetic problem: add the estimated duration of each activity, then subtract the total from the class's starting time.

The calculation also involves more than arithmetic. You are using probability, estimation, and decision-making under uncertainty. For example, if you take public transportation, your travel time depends partly on how frequently buses, jeepneys, or tricycles pass your stop. If vehicles arrive regularly, you may expect only a short waiting time. A long wait is possible, but you may consider it unlikely. Your estimate may also change depending on whether your class is held on a weekday or a Saturday, since traffic conditions are usually different.

You may also consider the possibility of rain, road construction, long queues, or unexpected delays. You might then add an extra fifteen or twenty minutes as a safety allowance. How large that allowance should be depends on how uncertain the journey is and how strongly you wish to avoid being late.

You have constructed a simple **mathematical model** of your morning. The model does not reproduce every detail of reality. Instead, it identifies the factors that matter most and uses them to guide a decision.

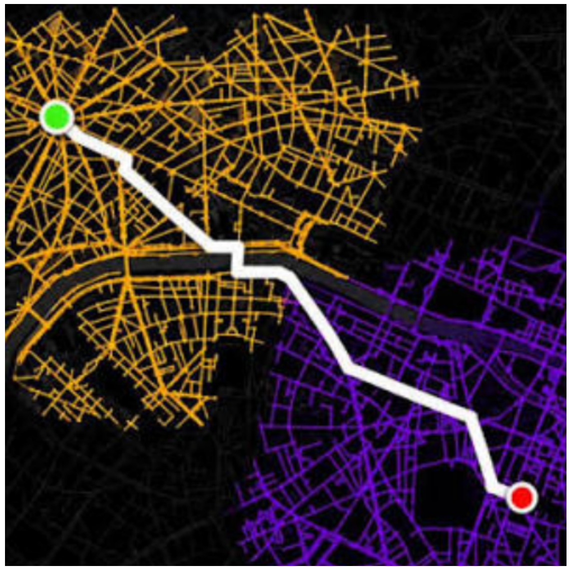


Figure 1.1: A conceptual illustration of shortest-path search across an urban road network represented as a weighted graph. The green and red markers denote the origin and destination, the white line shows the selected route, and the coloured regions indicate parts of the network explored during the search. Dijkstra's algorithm is a foundational graph-theoretic method for solving this kind of problem, although modern navigation applications employ more sophisticated variants.

Navigation applications such as Google Maps or Waze are familiar examples of systems organised around locations, routes, and travel-time estimates. At a simplified level, the mathematical toolkit for such systems includes:

- **Geometry and coordinate systems** represent locations, distances, and directions.
- **Graph theory** represents intersections as vertices and roads as weighted edges.
- **Optimisation algorithms**, including Dijkstra's algorithm [2] and A* [3], search for efficient routes through the resulting network.
- **Statistics and calculus** help estimate travel times and account for changing traffic conditions.

The estimated journey time comes from a mathematical model of location, networks, uncertainty, and change rather than from arithmetic alone.

Task

Watch the YouTube video *Why Navigation Apps Always Find the Perfect Route?* [4].

You will not be expected to implement Dijkstra's algorithm or A* as shown in the video. The point is to see how a familiar application uses mathematics behind the scenes. Real navigation systems extend these basic ideas to account for live traffic, road restrictions, and other practical conditions.

1.1 Patterns in Nature

Mathematics does more than help us make practical decisions. It also helps us recognise, describe, and investigate patterns—including patterns that people find beautiful.

Task

Watch the YouTube video *Nature by Numbers* [5].

1.2 The Golden Ratio

The video above brings together the famous Fibonacci sequence, the golden ratio, and several patterns associated with them. These ideas are closely connected, and the golden ratio provides the central thread linking much of the discussion.

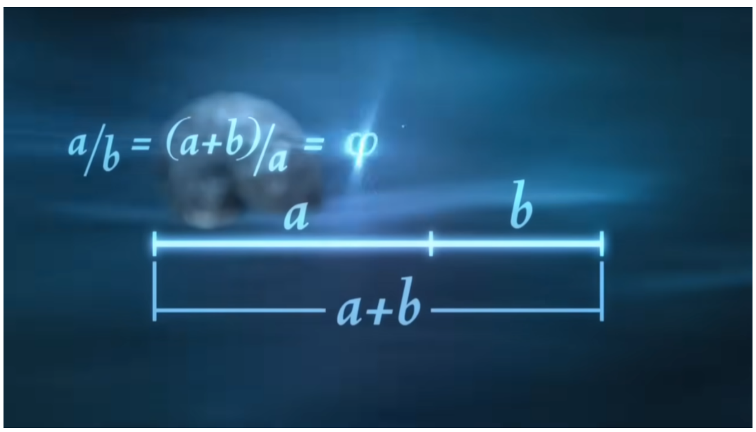
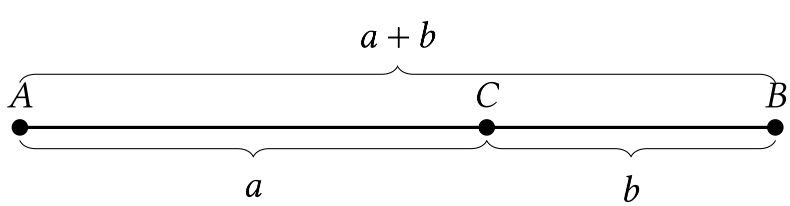


Figure 1.2: A frame from Nature by Numbers *illustrating the golden ratio [5].*

In Euclid's *Elements* [6], a straight line is said to be divided in the *golden ratio*[1] when

> *The whole is to the greater segment as the greater segment is to the lesser.*



Let's unpack this definition. Suppose a line segment AB is divided at a point C. Let the greater segment AC have length a and the lesser segment CB have length b. The whole segment AB then has length $a+b$. Euclid's definition states that the ratios ${}^{a+b}/_{a}$ and ${}^{a}/_{b}$ are equal. Their common value is the golden ratio, usually denoted by the Greek letter φ. Thus,

$$\varphi = \frac{a+b}{a} = \frac{a}{b}.$$

[1] Euclid calls this the *extreme and mean ratio*; here we use the modern term *golden ratio*.

This relationship allows us to calculate the numerical value of φ:

$$\varphi = \frac{a+b}{a} = 1 + \frac{b}{a} = 1 + \frac{1}{\varphi}.$$

Multiplying by φ gives $\varphi^2 - \varphi - 1 = 0$. The quadratic formula produces two roots. Because a ratio of lengths is positive, we retain the positive root:

$$\varphi = \frac{1+\sqrt{5}}{2} \approx 1.6180339887.$$

1.3 The Fibonacci Sequence

Fibonacci numbers arise in mathematical descriptions of several natural patterns, particularly some arrangements of leaves, petals, and seeds. Their appearance is not universal; when these patterns do occur, they can emerge from local rules governing growth and spatial arrangement [7].

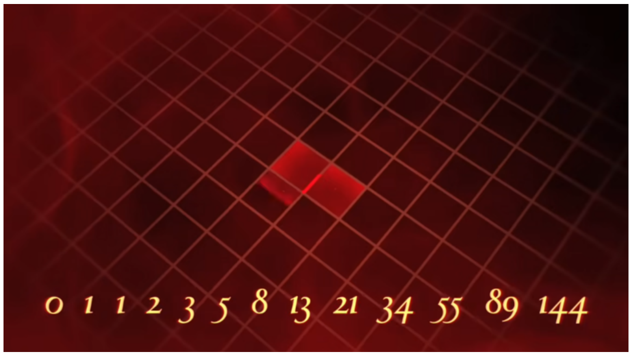


Figure 1.3: A frame from Nature by Numbers *illustrating the Fibonacci sequence [5].*

Spiral shells, including those of the nautilus, are often presented as examples of the Fibonacci sequence or golden ratio in nature [8]. A shell may resemble a logarithmic spiral, but resemblance alone does not show that it follows the Fibonacci sequence or golden ratio exactly.

The golden ratio and the Fibonacci sequence are closely related. Before examining that relationship, let us define the sequence formally.

The Fibonacci Sequence

The *Fibonacci sequence* is the sequence $F_0, F_1, F_2, F_3, \ldots$ defined by

$$F_0 = 0, \qquad F_1 = 1, \qquad F_n = F_{n-1} + F_{n-2} \quad \text{for } n \geq 2.$$

The first few terms are

$$\begin{aligned} F_0 &= 0 & F_3 &= F_2 + F_1 = 1 + 1 = 2, \\ F_1 &= 1 & F_4 &= F_3 + F_2 = 2 + 1 = 3, \\ F_2 &= F_1 + F_0 = 1 + 0 = 1 & F_5 &= F_4 + F_3 = 3 + 2 = 5. \end{aligned}$$

Continuing this process, the sequence begins 0, 1, 1, 2, 3, 5, 8, 13, 21, ... Each term after the first two is the sum of the two preceding terms. Now consider ratios of consecutive nonzero terms:

$$\frac{F_3}{F_2} = \frac{2}{1} = 2, \qquad \frac{F_4}{F_3} = \frac{3}{2} = 1.5, \qquad \frac{F_5}{F_4} = \frac{5}{3} \approx 1.667,$$

$$\frac{F_6}{F_5} = \frac{8}{5} = 1.6, \qquad \frac{F_7}{F_6} = \frac{13}{8} = 1.625, \qquad \frac{F_8}{F_7} = \frac{21}{13} \approx 1.615.$$

The ratios alternate above and below 1.618, drawing progressively closer to the golden ratio. This suggests a natural question: do the ratios actually approach φ?

From Pattern to Proof. Computing several ratios and using the pattern to form a conjecture is an example of **inductive reasoning**. The evidence suggests a general conclusion, but finitely many examples cannot prove a claim about the entire infinite sequence. **Deductive reasoning** instead begins with definitions, accepted facts, or clearly stated assumptions and uses valid logical steps to reach a conclusion. The distinction matters: a pattern may suggest what is true, while a proof explains why it must be true.

Fibonacci and the Golden Ratio (Non-examinable)

Recall that

$$F_n = F_{n-1} + F_{n-2}.$$

We are interested in the behaviour of F_n/F_{n-1} as n becomes large. Assume that this ratio approaches a positive limit L:

$$L = \lim_{n\to\infty} \frac{F_n}{F_{n-1}}.$$

Dividing the recurrence relation by F_{n-1} gives

$$\frac{F_n}{F_{n-1}} = 1 + \frac{F_{n-2}}{F_{n-1}} = 1 + \frac{1}{F_{n-1}/F_{n-2}}.$$

The shifted ratio F_{n-1}/F_{n-2} has the same limit L. Taking limits therefore gives

$$L = 1 + \frac{1}{L}.$$

Because $L > 0$, solving $L^2 - L - 1 = 0$ yields

$$L = \frac{1+\sqrt{5}}{2} = \varphi.$$

A complete proof must also establish that the limit exists; that additional step is beyond the scope of this course.[2]

[2] *(Non-examinable)* The argument above is only one route to the result. Other methods, including techniques from linear algebra, lead to the same conclusion. A correct and clearly explained method can receive full credit even if it

1.4 The Golden Angle and a Word of Caution

In botany, **phyllotaxis** describes how leaves, seeds, or other plant organs are arranged. In some spiral arrangements, successive organs form an angle close to the **golden angle**, $360^\circ/\varphi^2 \approx 137.5^\circ$. This spacing is associated with the familiar Fibonacci-like spirals seen in some flower heads and other plant structures.

Figure 1.4: A frame from Nature by Numbers *illustrating a golden-angle pattern in a sunflower head [5].*

This pattern is a genuine subject of research in mathematical biology [7], but it should not be treated as a universal rule. Not every plant follows a golden-angle pattern, and plants do not calculate φ. Rather, the pattern can emerge from biological growth processes and interactions among developing structures in limited space.

A word of caution. The golden ratio appears in some natural patterns, but claims about it are often exaggerated. Faces, paintings, buildings, shells, and galaxies are frequently described as "golden" without sufficient evidence.

A shape that looks approximately golden is not evidence that the ratio was deliberately used or that it caused the shape to form. A mathematical claim should be supported by clear definitions, reproducible measurements, and—when biology is involved—a plausible biological mechanism.

1.5 Mathematics and Biomedical Imaging

The golden ratio is an attractive example, but its practical importance is easily overstated. Biomedical imaging offers a more direct example of mathematics at work.

Figure 1.6 illustrates what a person might see with a healthy eye and with an eye affected by **diabetic retinopathy** (DR), a complication of diabetes. Diabetes can damage the small blood vessels in the retina, the light-sensitive layer at the back of the eye. In advanced cases, these vessels may leak or bleed, causing blurred vision, dark floating spots, or even blindness [13]. To screen

differs from the one presented in class. You are not expected to discover a more advanced approach, but you should feel free to use one if you can justify it.

Figure 1.5: The Parthenon in Greece [9]. Although the building is often said to have been designed around the golden ratio, historical and geometrical evidence does not support that claim [10–12].

Figure 1.6: An illustration of normal vision (left) and vision affected by advanced diabetic retinopathy (right). Early DR may cause no noticeable symptoms [13].

for the disease, clinicians examine the retina using **fundus photographs**—colour photographs of the back of the eye—such as those shown in Figure 1.7 [14].

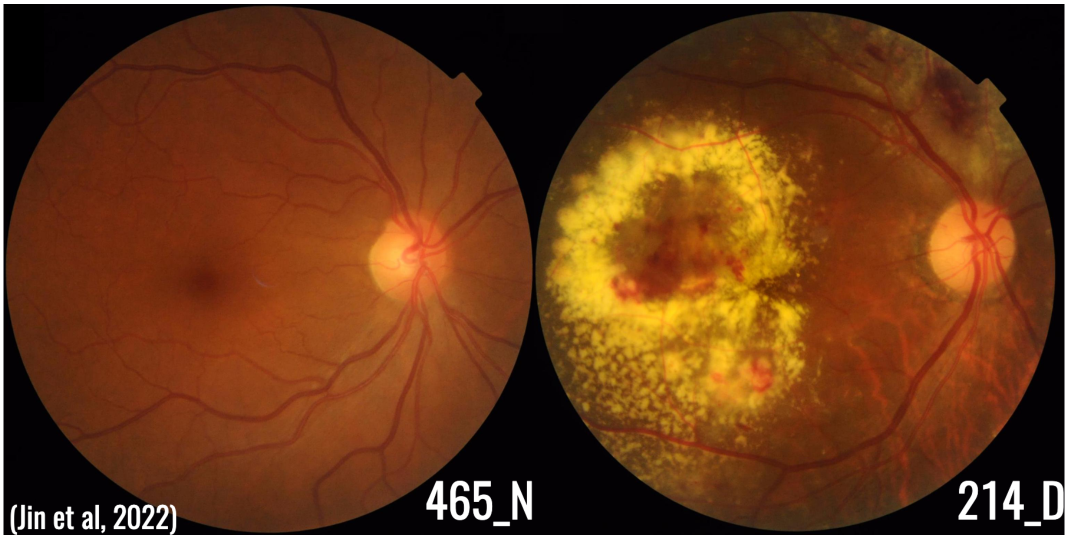


Figure 1.7: Fundus photographs from the FIVES dataset: a normal retina (left) and a retina affected by diabetic retinopathy (right) [15].

In this example, the image on the right appears much more irregular than the one on the left. It contains large yellow regions and other visible abnormalities. This particular case is relatively easy to distinguish by inspection, but not every case is so obvious. The signs of early disease can be subtle, and a diagnosis must always be made by a qualified clinician.

Some biology graduates go on to medicine and other health professions. Where many patients require screening, computers can organise images and flag those that merit closer review. These systems support clinical judgment; they do not replace it.

The diseased image in Figure 1.7 might be described informally as looking "messier". A computer cannot use such a description directly; it needs features that can be measured numerically.

One approach is **topological data analysis** (TDA), a mathematical framework that can be combined with machine learning and computer vision [16]. In my bachelor's thesis [17], I drew on ideas from several mathematical fields, including linear algebra, graph theory, algebraic topology, and category theory, to convert fundus images into numerical signatures. These signatures can then be supplied to a machine-learning system that classifies an image as either normal or affected by diabetic retinopathy. Figure 1.8 summarises this process.

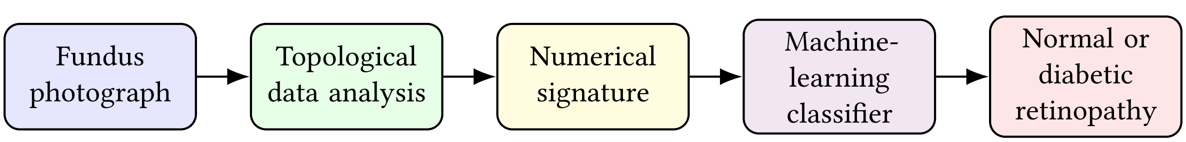


Figure 1.8: A simplified overview of the TDA-assisted retinal-image classification workflow [17].

Here, TDA records structural features in an image, such as separate components and loops, as numerical data [16]. In this retinal-image workflow, those data form a **topological signature** [17].

Once each image has a numerical signature, the signatures of different groups can be compared statistically. They can also be supplied to a machine-learning model, such as an artificial neural network.

This does not mean that TDA can diagnose diabetic retinopathy on its own. Digital images are already composed of numerical pixel values, and many computational methods analyse those values directly. TDA provides an additional description that focuses on shape and connectedness. Researchers must therefore test whether this additional information makes a computer-assisted screening system more accurate or reliable.

Machine Learning and Modern Artificial Intelligence. Machine learning provides much of the foundation for modern artificial intelligence, and many machine-learning systems employ **artificial neural networks**. Despite their biologically inspired name, these are mathematical and computational models rather than literal replicas of a brain.

Building and training these models draws on several disciplines. Computer science supplies the algorithms and software; linear algebra represents data and the connections within a network; calculus helps adjust those connections during training; and probability and statistics help describe uncertainty, detect patterns, and evaluate performance [18]. These systems depend on computation together with several branches of mathematics.

What the Retinal Example Shows

In this workflow, shapes and structures in a retinal image are converted to numbers, compared across groups, and examined for relationships with disease. Statistics and machine learning help account for uncertainty. The mathematics does not diagnose a patient; it gives the screening system another way to describe the image.

1.6 Is Mathematics an Exact Science?

Mathematics is often described as exact, but that statement needs an important qualification. **Pure mathematics can be exact within its definitions and assumptions.** Once we agree on the rules, a valid proof establishes its conclusion with logical certainty. The real world, however, does not arrive with perfectly defined objects, flawless measurements, or assumptions that are always true.

Reporting a Person's Height

Suppose someone reports a height of 5 feet 6 inches. This normally does not mean that the person's height is *exactly* 66 inches. The measurement may have been rounded to the nearest inch; it may also vary slightly with posture, footwear, and the measuring instrument. Nevertheless, "5 feet 6 inches" is usually precise enough for an ordinary conversation or a form asking for someone's height. Greater precision would add detail, but it would not necessarily be useful for the purpose at hand.

The same principle applies when mathematics is used to study traffic, populations, disease, climate, finance, or biological systems. A **mathematical model** keeps the features judged relevant to a question and leaves others out. It is therefore a simplified representation of reality, not reality itself. Its conclusions depend both on sound mathematics and on how well its assumptions and data represent the situation being studied.

Applied models can still be reliable for their intended purposes. To judge a model, ask what it is intended to do, how much accuracy the task requires, which assumptions it makes, and how uncertain its measurements and predictions are. A metro map, for example, is geographically distorted but remains useful for finding the correct train.

Exact Reasoning. Mathematics permits exact reasoning within a chosen model. In practice, the most useful model is often the simplest one accurate enough for the task.

1.7 Mathematics and Collaboration

Mathematicians are trained to work with quantities, patterns, structures, and carefully stated assumptions. That training does not supply all the biological knowledge needed to frame a problem well. As the familiar saying warns, "If all you have is a hammer, everything looks like a nail." A mathematician who approaches a biological problem using mathematics alone may construct an elegant model while overlooking something biologically important.

Effective applications therefore usually require collaboration with **domain experts**—people with specialised knowledge of the subject being investigated. In biology, these experts may include laboratory researchers, field biologists, clinicians, epidemiologists, and public-health specialists. They help determine which questions matter, which assumptions are reasonable, what data can be collected, and whether a mathematical conclusion makes biological sense.

Mathematical competence does not confer expertise in biology, medicine, or any other field to which mathematics is applied. A calculation can be correct even when the model behind it is biologically inappropriate. Biologists, in turn, benefit from enough mathematics to question a model's assumptions, interpret its results, and recognise its limitations. Good interdisciplinary work requires knowing when to ask questions, collaborate, or defer to someone with the relevant expertise.[3]

This division of work is common in **data science**, as Figure 1.9 illustrates. Domain experts frame and check the question, mathematicians and statisticians analyse data and uncertainty, and computer scientists build the systems needed to work with large or complex datasets [20].

A biologist need not become a mathematician to benefit from a working understanding of mathematical ideas. That understanding makes it easier to communicate with quantitative specialists and to recognise when a biological question can be measured, modelled, or investigated in a new way.[4]

Mathematics is one of the shared languages that allows specialists in computational, laboratory, clinical, and field work to tackle the same problem.

[3] *(Non-examinable)* Philosophers use the term **epistemic trespassing** for passing judgment on matters outside one's field of expertise. Awareness of this risk encourages intellectual humility: expertise in one discipline should not be assumed to transfer automatically to another [19].

[4] My academic background is primarily in mathematics and computer science. For the biological parts of my thesis, I relied on established work by biologists. I could not have framed or interpreted the problem from mathematics alone.

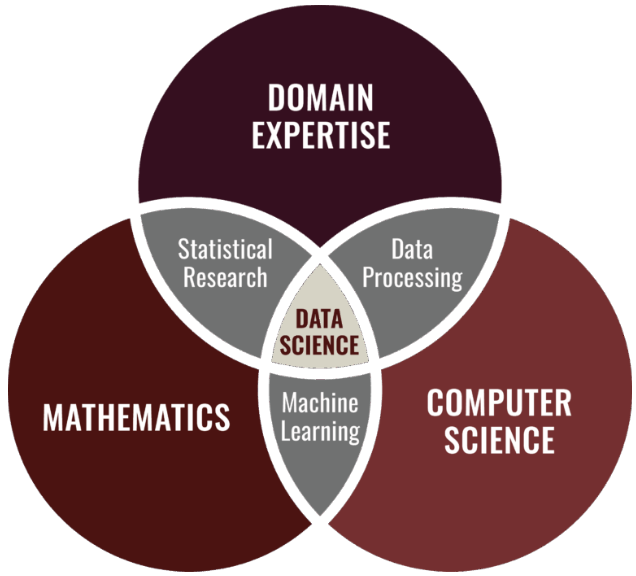


Figure 1.9: A common view of data science as the intersection of domain expertise, mathematics, and computer science. The overlapping areas also support statistical research, data processing, and machine learning [20].

1.8 What It Means to Think Mathematically

To think mathematically is to decide what information matters, identify patterns and relationships, state assumptions, build and test models, account for uncertainty, and explain why a conclusion follows. Calculation is one tool in that work.

1.9 Exercises

The questions move from everyday modelling to biological imaging. Explain the limits of each mathematical claim you make.

1. List the factors that could affect how long it takes you to travel to class. Which factors would you include in a useful model, and which details could you reasonably leave out?
2. Calculate the next three Fibonacci numbers and the ratios of successive terms. Do the new ratios support the conjecture that they approach φ? Explain why this is evidence rather than a complete proof.
3. Compare the two retinal photographs above. Suggest one visible feature that could be measured numerically. Explain what information your proposed measurement captures and what it might overlook.
4. Find an everyday or biological claim involving a mathematical pattern. What definition, measurements, or evidence would you require before accepting it?
5. A classmate says that mathematics is merely a collection of formulas for obtaining numerical answers. Write a short response challenging this view. Use two examples from the chapter to

explain how mathematics can be expressed, represented, and used, and identify at least one practical, intellectual, or aesthetic dimension of the subject.

6. Suppose the developers of a computer-assisted retinal-screening system know that its performance deteriorates when photographs are blurry, but advertise only its overall accuracy. What information should they disclose? Explain how honesty, integrity, and awareness of a model's limitations should affect the use of such a system in healthcare.

Closing Reflection

What decisions have you already made today that involved mathematics, even though you did not consciously perform a written calculation?

Chapter 2

The Language of Mathematics

Unfamiliar notation can make mathematics appear more intimidating than it really is. Mathematical symbols are not meant to conceal ideas. Rather, they allow us to express those ideas efficiently and with less ambiguity than ordinary language sometimes permits.

Learning Outcomes

After completing this chapter, you should be able to:

- distinguish among mathematical expressions, open sentences, and closed sentences, and translate elementary statements between words and symbols;
- read common mathematical symbols and apply standard conventions, including the order of operations;
- use sets, ordered pairs, relations, functions, and binary operations in elementary examples;
- interpret elementary logical connectives and quantifiers, distinguish a conditional statement from its converse, and negate simple quantified statements;
- explain how formality reduces ambiguity in mathematical communication; and
- recognise these ideas in biology and everyday life.

Like any language, mathematics has a vocabulary, a grammar, and a set of conventions. Its vocabulary consists of symbols and specialised terms. Its grammar determines which arrangements of symbols are meaningful. Its conventions allow different readers to interpret the same notation consistently.

Task

Watch the YouTube video *The Language of Maths* [21].[1]

Once its symbols and terms are defined, mathematical notation can state relationships precisely and in little space. It can also make complicated structures easier to analyse. It still needs prose: symbols record relationships compactly, while sentences explain what they mean and why they matter.

[1] *(Non-examinable)* In American English, *mathematics* is usually shortened to *math*. In British and Australian English, it is usually shortened to *maths*; hence the title of the video.

Example: Translating a Temperature Condition. Consider the sentence "The body temperature T is at least 38°C." The phrase **at least** includes 38, so we write $T \geq 38$.

The symbolic form is **concise**: it expresses the condition using only a few symbols. It is also **precise** once we have established that T denotes temperature measured in degrees Celsius and that $\geq$ means "is greater than or equal to." In contrast, "the body temperature is greater than 38°C" is written as $T > 38$. Removing the small bar beneath $>$ excludes a temperature of exactly 38°C.

The symbolic condition can also be compared with other inequalities and manipulated using algebraic rules.

2.1 Expressions and Sentences

The distinction between a mathematical **expression** and a mathematical **sentence** is similar to the distinction between a noun phrase and a complete sentence in English.

Expressions and Sentences

A **mathematical expression** is a meaningful arrangement of symbols that represents a mathematical object, value, or quantity. An expression does not, by itself, make a claim that can be judged true or false.

A **mathematical sentence** is a meaningful arrangement of symbols that makes a mathematical claim. A **closed sentence**, also called a **logical proposition**, can be judged either true or false.

An assertion whose truth depends on the value assigned to an unspecified variable is called an **open sentence**. Assigning a value to the variable, or introducing a phrase such as "for every" or "there exists," can turn an open sentence into a closed sentence.

A useful first test is to read the notation aloud. A mathematical sentence makes a complete claim, whereas an expression does not. A sentence is open when one or more of its variables still need values before the claim can be judged true or false. It is closed when no such assignment is needed. The appearance of a variable does not automatically make a sentence open. In $\forall x \in \mathbb{R},\ x^2 \geq 0$, the phrase "for every x" tells us how x is to be interpreted, so the statement can be judged true or false as written.[2]

Mathematical sentences often join expressions using relation symbols such as $=$, $\neq$, $<$, $\leq$, and $\in$. In the sentence $1 + 1 = 2$, for example, the equality sign asserts that the expressions on its two sides represent the same number.

2.1.1 Reading Common Mathematical Symbols

Symbols should be read as meaningful words or phrases, not as decorative marks. The following table lists some symbols that will appear throughout these notes.

[2] *(Non-examinable)* Expressions, sentences, variables, and quantifiers are studied formally in **mathematical logic**. Much of the notation used here belongs to **first-order predicate logic**, also called **first-order predicate calculus**. A **Hilbert-style calculus** is one formal system for constructing proofs in logic [22]. Here, *calculus* means a systematic method of calculation or reasoning, not necessarily differentiation and integration. Differential and integral calculus became so influential that they nearly monopolised the word, but many other calculi exist. The plural of *calculus* is *calculi*.

Notation	Classification	Reason
$b^2 - 4ac$	expression	represents the discriminant of $ax^2 + bx + c$ but makes no assertion
$\dfrac{y_2 - y_1}{x_2 - x_1}$	expression	may represent the slope of a line but makes no assertion
$x + 1 = 2$	open sentence	its truth depends on the value assigned to x
$2x + 3 > 7$	open sentence	its truth depends on the value assigned to x
$1 + 1 = 2$	true closed sentence	it makes a true assertion and contains no unspecified variables
$1 + 1 = 3$	false closed sentence	it makes a false assertion and contains no unspecified variables
$\forall x \in \mathbb{R},\ x^2 \geq 0$	true closed sentence	the variable is governed by "for every", and every real number has a nonnegative square
$\exists x \in \mathbb{R},\ x^2 = -1$	false closed sentence	no real number has a negative square

Symbol	Meaning	Read as	Example
$=$	equality	equals	$2 + 3 = 5$
$\neq$	non-equality	is not equal to	$2 + 3 \neq 6$
$<$	strict inequality	is less than	$3 < 5$
$>$	strict inequality	is greater than	$8 > 4$
$\leq$	inequality	is less than or equal to	$x \leq 10$
$\geq$	inequality	is greater than or equal to	$x \geq 0$
$\approx$	approximate equality	is approximately equal to	$\pi \approx 3.14$
$\pm$	two possible signs	plus or minus	$x = \pm 2$
$\in$	set membership	is an element of	$3 \in \mathbb{N}$
$\notin$	set non-membership	is not an element of	$-1 \notin \mathbb{N}$
$\mathbb{N}$	natural numbers	the natural numbers	$5 \in \mathbb{N}$
$\mathbb{Z}$	integers	the integers	$-3 \in \mathbb{Z}$
$\mathbb{Q}$	rational numbers	the rational numbers	$\frac{3}{4} \in \mathbb{Q}$
$\mathbb{R}$	real numbers	the real numbers	$\sqrt{2} \in \mathbb{R}$
$\mathbb{C}$	complex numbers	the complex numbers	$2 + 3i \in \mathbb{C}$
$\subseteq$	subset	is a subset of	$\mathbb{N} \subseteq \mathbb{Z}$
$\cup$	union of sets	union	$A \cup B$
$\cap$	intersection of sets	intersection	$A \cap B$
$\forall$	universal quantifier	for every	$\forall x \in \mathbb{R},\ x^2 \geq 0$
$\exists$	existential quantifier	there exists	$\exists x \in \mathbb{R},\ x^2 = 4$
$\Rightarrow$	implication	implies	$x > 2 \Rightarrow x > 0$
$\Leftrightarrow$	equivalence	if and only if	$x = 0 \Leftrightarrow x^2 = 0$
$\therefore$	conclusion	therefore	$x = 2,\ \therefore x^2 = 4$
$:=$	definition	is defined as	$f(x) := x^2 + 1$
$\sum$	summation	the sum of	$\sum_{i=1}^{n} x_i$
$\prod$	repeated multiplication	the product of	$\prod_{i=1}^{n} x_i$
$\propto$	proportionality	is proportional to	$y \propto x$

Do I Need to Memorise All These Symbols? No. You may write a mathematical statement in words, provided that its meaning is clear and correct. Learning the common symbols is nevertheless worthwhile: they make mathematical writing shorter, save time when you write solutions, and help you read textbooks and scientific material more easily. Familiarity will develop through use, so you do not need to master every symbol at once. This compactness is also one reason why some mathematics textbooks are physically thinner than textbooks in subjects such as biology: a short line of notation can express what might otherwise require several sentences. If every symbol and intermediate step were repeatedly written out in full, the same mathematical material could occupy many more pages.

2.1.2 Translating Between Words and Symbols

Translating between ordinary language and mathematical notation requires attention to meaning rather than word-for-symbol substitution.

Additional Translation Examples.

Words	Symbols
x is positive	$x > 0$
n is no more than 20	$n \leq 20$
p is between 0 and 1, inclusive	$0 \leq p \leq 1$
Population density d is population N divided by area A	$d = \dfrac{N}{A}$
x belongs to the set A	$x \in A$
A and B have no common elements	$A \cap B = \emptyset$
There is a real number whose square is 4	$\exists x \in \mathbb{R} : \ x^2 = 4$

For example, the order of words matters when translating the phrase “less than”. The phrase “five less than x” means “$x - 5$”, and not “$5 - x$”.

2.2 Conventions in Mathematical Language

Mathematical conventions are agreed ways of writing and interpreting mathematics. Without them, two readers could assign different meanings to the same line of notation.

2.2.1 Order of Operations

One familiar convention is the order of operations, often remembered using **PEMDAS**: parentheses, exponents, multiplication, division, addition, and subtraction.

Multiplication and division have equal priority, as do addition and subtraction. Operations of equal priority are performed from left to right.

Example: Applying the Order of Operations. Consider

$$\begin{aligned} 1 + 2 - 3(4 + 5) &= 1 + 2 - 3(9) && \text{parentheses} \\ &= 1 + 2 - 27 && \text{multiplication} \\ &= 3 - 27 && \text{addition} \\ &= -24. && \text{subtraction} \end{aligned}$$

Following the agreed convention gives a unique result. Performing the subtraction before the multiplication would produce a different answer.

2.2.2 Other Common Conventions

- If n is a positive integer, then $x^n = \underbrace{x \cdot x \cdots x}_{n \text{ factors}}$.
- Symbols written next to one another usually indicate multiplication: $3x = 3 \cdot x$ and $ab = a \cdot b$.
- A fraction bar groups the entire numerator and denominator: $\dfrac{a+b}{c+d} = (a+b) \div (c+d)$.
- Parentheses can change the quantity to which an operation applies: $-3^2 = -(3^2) = -9$, whereas $(-3)^2 = 9$.
- The notation $f(x)$ usually denotes the value of the function f at x. It does not ordinarily mean f multiplied by x.
- The equality sign expresses sameness of value. Every expression in a chain such as $a = b = c$ must represent the same object or value.
- The approximation sign $\approx$ should be used when two quantities are close but not exactly equal. For example, $\pi \approx 3.14$, not $\pi = 3.14$.

Conventions are not laws of nature. They are shared agreements that make mathematical communication efficient and consistent.

2.3 Sets, Relations, Functions, and Binary Operations

Sets collect objects; relations describe connections among them; functions assign outputs to inputs; and binary operations combine two objects to produce a third. We begin with sets.

2.3.1 Sets

For this course, the following informal definition is sufficient.[3]

Set

For the purposes of this course, a **set** is a well-defined collection of distinct objects. The objects belonging to the set are called its **elements**.

Set Notation. Sets are usually denoted by capital letters. For example, let $D = \{A, C, G, T\}$, where A, C, G, and T denote adenine, cytosine, guanine, and thymine, respectively. Then $A \in D$, meaning that adenine is an element of D.

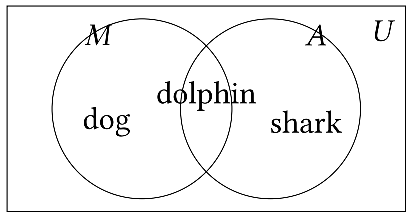
Repeating an element does not change a set, and the order of the elements does not matter. Thus, $\{1, 2, 2, 3\} = \{3, 2, 1\} = \{1, 2, 3\}$.

If every element of a set A is also an element of a set B, we write $A \subseteq B$ and say that A is a **subset** of B. The set containing no elements is called the **empty set** and is denoted by $\varnothing$.

[3] *(Non-examinable)* This is the usual informal definition from **naïve set theory**. In more advanced mathematics, allowing any verbal description to define a set can produce contradictions, including Russell's paradox [23]. Modern axiomatic set theory restricts how sets may be constructed in order to avoid such contradictions.

Set Operations. Two important operations on sets are union and intersection. The **union** $A \cup B$ contains the elements that belong to A, to B, or to both. The **intersection** $A \cap B$ contains only the elements common to both sets.

Example: A Venn Diagram. A Venn diagram represents sets as regions. Let the universal set U contain a dog, a dolphin, and a shark. Let M contain the mammals in U, and let A contain the aquatic animals. The dolphin belongs to both sets, so it is placed in their overlap.



Thus, $M \cap A = \{\text{dolphin}\}$ and $M \cup A = \{\text{dog}, \text{dolphin}, \text{shark}\}$.

2.3.2 Ordered Pairs and Cartesian Products

Example: Ordered Pairs and the Cartesian Plane. An **ordered pair** (a, b) records both its entries and their order; in general, $(a, b) \neq (b, a)$.

The **Cartesian product** of two sets A and B is

$$A \times B = \{(a, b) : \ a \in A \text{ and } b \in B\}.$$

It is the set of all ordered pairs whose first entry belongs to A and whose second entry belongs to B.

The familiar Cartesian plane is $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. Each point (x, y) represents one location: x gives its horizontal position and y gives its vertical position.

2.3.3 Relations

Relation

A **relation from** A **to** B is a subset of $A \times B$. It records which elements of A are related to which elements of B.

Relations may also be defined on a single set. In an ecological network, for example, "preys upon" is a relation on a set of species.

Example: A Relation as a Dietary Model. Let A contain a dog, a wolf, and a hyena, and let $B = \{\text{animal matter}, \text{plant matter}\}$. In a deliberately simplified model, consider the relation

$$\begin{aligned} R = \{&(\text{dog}, \text{animal matter}), (\text{dog}, \text{plant matter}), \\ &(\text{wolf}, \text{animal matter}), (\text{hyena}, \text{animal matter})\}. \end{aligned}$$

Here, the dog is paired with two elements of B.

Now let $B' = \{\text{carnivore}, \text{herbivore}, \text{omnivore}\}$ and define

$$R' = \{(\text{dog}, \text{omnivore}), (\text{wolf}, \text{carnivore}), (\text{hyena}, \text{carnivore})\}.$$

Unlike R, the relation R' pairs every element of A with exactly one element of B'. Relations with this property form an important special class called **functions**, which we define in the next section.

Both relations deliberately simplify real animal diets. They are useful here because they make the difference between R and R' easy to see.

Example: Family Relationships as a Relation. Let P be a set of people. We can define a relation R on P by declaring that aRb if and only if a is a relative of b. Thus, $(a, b) \in R$ means that person a is a relative of person b.

Example: Number Systems as a Relation. Let $A = \mathbb{C}$ and $B = \{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}\}$. Define $x\,R\,S$ to mean that $x \in S$. For example, $2\,R\,\mathbb{N}$, $2\,R\,\mathbb{Z}$, and $2\,R\,\mathbb{R}$, while $-3\,R\,\mathbb{Z}$ and $\sqrt{2}\,R\,\mathbb{R}$.

2.3.4 Functions

Functions are a special type of relation.

Function

A **function** f from a set A to a set B, written $f\colon A \to B$, is a relation that assigns **every** element of A to **exactly one** element of B.

The set A is called the **domain**, and B is called the **codomain**. The set of outputs actually produced by the function is its **range**. The range is a subset of the codomain.

The phrase "exactly one" is essential. Different inputs may have the same output, but a single input cannot be assigned two different outputs by the same function. A function $f\colon A \to B$ must also assign an output to every element of its stated domain A.

Example: DNA Base Pairing as a Function. Let $D = \{A, C, G, T\}$, where the letters denote adenine, cytosine, guanine, and thymine, respectively. Figure 2.1 shows these bases within the familiar double-helix structure.

Figure 2.1: An illustration of the DNA double helix. Its complementary base-pairing rules can be modelled as a function [24].

Complementary base pairing pairs adenine with thymine and cytosine with guanine. This rule defines a function $f\colon D \to D$ by

$$f(A) = T, \qquad f(T) = A, \qquad f(C) = G, \qquad f(G) = C.$$

Every input base has exactly one complementary base, so this rule is a function.

2.3.5 Binary Operations

Binary Operation

A **binary operation** on a set A is a function $\star\colon A \times A \to A$. It takes two elements of A, combines them according to a specified rule, and produces another element of A.

Examples of Binary Operations. **Addition.** Addition is a binary operation on the integers, $+\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, because the sum of any two integers is another integer.

Set union. Similarly, if $\mathscr{P}(U)$ is the set of all subsets of a fixed set U, then union is a binary operation $\cup\colon \mathscr{P}(U) \times \mathscr{P}(U) \to \mathscr{P}(U)$.

Example: When an Operation Is Not Binary. The underlying set matters. Subtraction is not a binary operation on the positive integers because $2 - 3 = -1$, which is not positive. Division is not a binary operation on $\mathbb{R}$ because division by zero is undefined.

Example: A Custom Binary Operation. The meaning of an operation symbol comes from its definition. An unfamiliar symbol need not represent one of the familiar arithmetic operations.

For example, define an operation $\star$ on the real numbers by $a \star b = a + 2b$. Applying the stated rule gives $3 \star 4 = 3 + 2(4) = 11$.

2.4 Elementary Logic

Sets, relations, functions, and operations describe mathematical objects and how they are connected. Logic describes the claims we make about those objects.

Logic studies valid reasoning: how statements are combined and which conclusions follow from them.

Scope of this section. We need enough logic to read the basic connectives and quantifiers, interpret simple statements, complete a two-proposition truth table, and distinguish a conditional from its converse. Longer truth tables and formal proofs of logical equivalence lie beyond our scope.

Logical Proposition

A **logical proposition** is a closed sentence that is either true or false. Under a fixed interpretation, it cannot be both true and false.[4]

For example, "A dog is a mammal" is a logical proposition. The expression $x + 1$ makes no claim, while the open sentence $x + 1 = 2$ cannot be judged true or false until x is specified or quantified.

2.4.1 Logical Connectives

Simple logical propositions can be joined to form compound logical propositions. Let P and Q denote logical propositions.

Symbol	Name	Reading	When it is true
$\neg P$	negation	not P	when P is false
$P \wedge Q$	conjunction	P and Q	when both P and Q are true
$P \vee Q$	disjunction	P or Q	when at least one of P and Q is true
$P \Rightarrow Q$	conditional	if P, then Q	except when P is true and Q is false
$P \Leftrightarrow Q$	biconditional	P if and only if Q	when P and Q have the same truth value

In mathematics, the word **or** is normally inclusive. Thus, $P \vee Q$ is true when P is true, when Q is true, or when both are true.

Example: A Truth Table for the Logical Connectives. A **truth table** displays the truth value of a compound logical proposition for every possible combination of truth values of its components.

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

Example: A Conditional and Its Converse. Let P mean "the organism is a dog" and Q mean "the organism is a mammal." The conditional $P \Rightarrow Q$ says, "If the organism is a dog, then it is a mammal."

Under the usual biological classification, this statement is true. Its **converse**, $Q \Rightarrow P$, says, "If the organism is a mammal, then it is a dog."

The converse is false because many mammals are not dogs. Thus, a conditional statement and its converse need not have the same truth value.

[4]*(Non-examinable)* In this section, **logical proposition** means a statement that can be judged true or false. Mathematical writers also use **Proposition** as a label for a proved result. A theorem is usually presented as a major result; a proposition as a useful stand-alone result that the author considers less central; and a lemma as a result used to prove a later theorem or proposition. These labels are not strict rankings and may vary from one text to another. Theorems, propositions, and lemmas are all logical propositions because they are statements that can be judged true or false.

Implication Is Not Necessarily Causation. The conditional $P \Rightarrow Q$ is false only when P is true and Q is false; in every other case it is true. This truth-table rule says nothing about causation.

For example, consider a particular accident. Let P mean "I banged my head against a wall" and let Q mean "my head hurts." If both propositions are true, then the conditional $P \Rightarrow Q$ is true. The notation alone, however, does not state that the impact caused the pain. Now let R mean "I have a bump on my forehead." The conditional $R \Rightarrow Q$ is also true when both propositions are true, but it does not follow that the bump caused the pain; the impact may have caused both. A true conditional may accompany a causal relationship, but it does not establish one. Nor does $P \Rightarrow Q$ imply that the converse $Q \Rightarrow P$ is true.

2.4.2 Quantifiers

Quantifiers state whether a condition holds for every element, for at least one element, or for exactly one element.

Symbol	Reading
$\forall$	for every, or for all
$\exists$	there exists at least one
$\exists!$	there exists exactly one

Example: Quantifying DNA Base Pairing. The DNA complementary-base function satisfies $\forall x \in D,\ f(f(x)) = x$. This says that, for every DNA base x in D, taking the complement twice returns the original base.

Example: Why the Domain Matters. The domain of a quantifier matters. Consider the logical proposition

$$\forall x \in \mathbb{Z}, \quad \exists y \in \mathbb{Z} \text{ such that } x + y = 0.$$

This proposition is true. Given any integer x, we may choose $y = -x$, which is also an integer. The number y is called the **additive inverse** of x.

By contrast, the proposition

$$\forall x \in \mathbb{N}, \quad \exists y \in \mathbb{N} \text{ such that } x + y = 0$$

is false when $\mathbb{N}$ denotes the natural or counting numbers. For example, if $x = 1$, the equation forces $y = -1$, but $-1 \notin \mathbb{N}$.

The same equation can therefore lead to different truth values when the permitted domain changes.

Example: Biological Classification Depends on the Domain. The same issue arises in biological classification. Consider the traditional five-kingdom model, and let $\mathscr{L}$ be the set of living cellular organisms covered by that model [25]. Let

$$\mathscr{K} = \{\text{Monera}, \text{Protista}, \text{Fungi}, \text{Plantae}, \text{Animalia}\},$$

where each kingdom is itself treated as a set of organisms. The claim that every organism receives a kingdom classification can be expressed as

$$\forall x \in \mathscr{L}, \quad \exists Y \in \mathscr{K} \text{ such that } x \in Y.$$

Now suppose we broaden the domain to a set $\mathscr{G}$ containing all biological entities that possess a genome, while leaving $\mathscr{K}$ unchanged. The corresponding proposition

$$\forall x \in \mathscr{G}, \quad \exists Y \in \mathscr{K} \text{ such that } x \in Y$$

is false because the enlarged domain includes viruses, which are not assigned to any of the five kingdoms.

The varicella-zoster virus—which causes chickenpox and shingles—provides one counterexample. It has a double-stranded DNA genome [26], but it does not belong to Monera, Protista, Fungi, Plantae, or Animalia.

Defining the domain as "entities that possess a genome" also allows us to avoid the separate and contested question of whether viruses should be regarded as living.

Changing the domain exposes a limitation of the five-kingdom model. The quantified proposition is true for the organisms the classification was designed to cover but false after viruses are included.

Example: Negating Quantified Propositions. Negating a quantified proposition changes its quantifier:

$$\neg(\forall x,\ P(x)) \iff \exists x \text{ such that } \neg P(x), \qquad \neg(\exists x,\ P(x)) \iff \forall x,\ \neg P(x).$$

In words, the negation of "Every organism has property P" is not "No organism has property P." The correct negation is the weaker statement "At least one organism does not have property P."

Logical notation makes the structure of an argument explicit.

Task

Watch *An Introduction to Propositional Logic* [27].

2.5 Formality in Mathematical Language

In ordinary conversation, context, tone, and shared experience often help us interpret incomplete or ambiguous statements. Mathematics aims to reduce this dependence on context.

Formality means expressing an idea with enough structure that its objects, assumptions, and logical relationships can be interpreted consistently.

Example: Formalising a Classification Rule. Consider the informal sentence "Each animal has one classification." This statement does not identify the animals or the possible classifications. Moreover, the word "one" might mean "at least one" or "exactly one".

In the dietary model, the sets A and B' and the relation R' allow us to write the intended claim as

$$\forall a \in A, \qquad \exists! b \in B' \text{ such that } (a, b) \in R'.$$

This says that every animal receives exactly one classification. The words define the objects, while the notation makes the logical structure visible.

Making a Statement Sufficiently Formal. Before presenting a mathematical claim, identify the objects being discussed, define the variables and the values they may take, state any necessary assumptions, use the required quantifiers and logical connectives, and make clear exactly what is being asserted. The appropriate degree of formality depends on the purpose and audience. A classroom explanation may use more ordinary language than a formal proof, but neither should rely on avoidable ambiguity.

2.6 Exercises

Classify, translate, calculate, and explain. Where a statement is ambiguous, say what information is missing.

1. Classify each item as an expression, an open sentence, or a closed sentence:

   a. $3x + 5$
   b. $3x + 5 = 20$
   c. $3(5) + 5 = 20$
   d. $\forall x \in \mathbb{R},\ x^2 \geq 0$

2. Translate each item as directed:

   a. The population P is greater than 500.
   b. The probability p is between 0 and 1, inclusive.
   c. Species s belongs to the set E of endangered species.
   d. Write $A \cap B = \emptyset$ in words.
   e. Write $\forall x \in D,\ f(f(x)) = x$ in words.
   f. The average population growth rate r is the change from N_1 to N_2 divided by the elapsed time from t_1 to t_2.

3. Evaluate the following expression and explain which convention you use at each step:

$$20 - 12 \div 3 \times 2.$$

4. Let

$$U = \{\text{bat}, \text{eagle}, \text{penguin}, \text{whale}\},$$

   and let M contain the mammals in U, while F contains the animals in U that can fly. Draw a Venn diagram for M and F, then state $M \cap F$ and $M \cup F$.

5. Let

$$A = \{\text{cat}, \text{dog}, \text{eagle}\}, \qquad B = \{\text{mammal}, \text{bird}\}.$$

Write a relation from A to B that records the group to which each animal belongs. Is this relation also a function? Explain.

6. Complete the following truth table. Then let P mean "the organism is a dog" and Q mean "the organism is a mammal." Write $P \Rightarrow Q$ and its converse in words.

P	Q	$\neg Q$	$P \wedge \neg Q$
T	T		
T	F		
F	T		
F	F		

7. Write the negation of each statement without simply placing the words "it is not true that" at the beginning:
   a. Every organism in the sample exhibits the trait.
   b. There exists a patient whose temperature satisfies $T \geq 38^\circ\text{C}$.

8. Decide whether each of the following is a binary operation on the stated set:
   a. addition on $\mathbb{Z}$;
   b. subtraction on the positive integers;
   c. multiplication on $\mathbb{R}$; and
   d. division on $\mathbb{R}$.

9. On $\mathbb{R}$, let $a \star b = a + 2b$. Calculate

$$5 \star (-2), \qquad (1 \star 3) \star 2, \qquad 1 \star (3 \star 2).$$

   Is $\star$ associative?

10. Give a biological example of a set, a relation, and a function.

11. The sentence "Each specimen has one measurement" is ambiguous. List three pieces of information needed to turn it into a precise mathematical claim.

Closing Reflection

Choose one mathematical symbol that once seemed unfamiliar to you. Explain its meaning in an ordinary English sentence and give an example of its use.

Chapter 3

Problem-Solving Strategies and Formulas

An unfamiliar problem rarely announces which calculation or formula to use. We must identify what is being asked, choose an approach, carry it out, and check the result. George Pólya organised this work into four broad stages. We begin with those stages before using them to understand familiar formulas.

Learning Outcomes

After completing this chapter, you should be able to:

- describe the four stages of Pólya's approach to problem solving;
- apply those stages to an accessible mathematical problem;
- explain why reformulating a problem may reveal a solution;
- follow a derivation of a familiar formula;
- derive and apply the sum formula for an arithmetic series;
- distinguish memorising a formula from understanding its assumptions; and
- organise a solution so that another reader can follow and assess it.

3.1 Pólya's Four Steps

Pólya's stages are not a rigid recipe: a solver may move backwards and forwards between them. They nevertheless provide a useful structure for approaching an unfamiliar problem. Suppose, for example, that we want to find the sum $1 + 2 + \cdots + 100$.

1. **Understand the problem.** Identify what is given, what must be determined, and what the notation means. Here the ellipsis denotes every positive integer from 1 to 100, not merely the three displayed terms.

2. **Devise a plan.** Connect the unfamiliar problem to something already known or change its representation. Pair the first term with the last, the second with the second-to-last, and so on. Each pair has the same sum: $1 + 100 = 101$, $2 + 99 = 101$, and $3 + 98 = 101$.

3. **Carry out the plan.** Perform each step carefully and justify it. There are 50 pairs, each with sum 101, so the required sum is $50(101) = 5050$.

4. **Look back.** Check whether the result answers the original question and agrees with the conditions of the problem. The average of the first and last terms is 50.5; multiplying by 100 terms also gives 5050, which checks the result in a different way.

Intuition may suggest a direction. Justification tells us whether that direction works.

3.2 Deriving and Using a Formula

Pólya's stages can guide the discovery of a general formula as well as the solution of a particular problem. The quadratic formula is a familiar example. Suppose that $a, b, c \in \mathbb{R}$, with $a \neq 0$, and $ax^2 + bx + c = 0$. We want to express its real solutions in terms of a, b, and c. Completing the square provides the plan.

Because $a \neq 0$, divide by a and move the constant term to the right-hand side:

$$x^2 + \frac{b}{a}x = -\frac{c}{a}.$$

Adding $\left(\frac{b}{2a}\right)^2$ to both sides gives

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a},$$

and hence

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}.$$

Equivalently,

$$(2ax + b)^2 = b^2 - 4ac.$$

If $b^2 - 4ac < 0$, the equation has no real solutions. Otherwise, taking square roots with both signs and solving for x gives

$$\boxed{x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}.$$

This is the **quadratic formula**. The derivation preserves the solution set: it divides by a known nonzero number, performs the same operation on both sides, and includes both signs when square roots are taken.

The quantity $\Delta = b^2 - 4ac$ is the **discriminant**. If $\Delta < 0$, the equation has no real solutions; if $\Delta = 0$, it has one repeated real solution; and if $\Delta > 0$, it has two distinct real solutions.

We can now use the discriminant without deriving the formula again.

Applying the discriminant. Consider the equation $x^2 + x + 1 = 0$. Here, $a = 1$, $b = 1$, and $c = 1$, so $b^2 - 4ac = 1^2 - 4(1)(1) = -3$. In the quadratic formula, the radical $\sqrt{-3}$ is not real; therefore, the equation has no real solutions.

3.3 Formulas Are Not Magic

Students often first meet the quadratic formula as something to memorise. That is a practical beginning, not its mathematical origin: the formula follows by completing the square in the general quadratic equation. A student may learn to use it before having enough algebra to follow every step of the derivation, but the derivation is still there. Formulas record relationships; they are not arbitrary strings of symbols.

3.3.1 An Arithmetic Sequence and Its Sum

Arithmetic Sequence

An **arithmetic sequence** is a list of numbers in which the difference between consecutive terms is constant.

An arithmetic sequence and its series. The sequence $2, 5, 8, 11, 14$ is arithmetic, with first term 2 and common difference 3. The expression obtained by adding its terms, $2 + 5 + 8 + 11 + 14$, is the corresponding **arithmetic series**.

Suppose an arithmetic sequence has n terms. Let t_1 denote its first term, d its common difference, and t_n its final term. Then

$$t_n = t_1 + (n-1)d.$$

Its associated arithmetic series is

$$S = t_1 + (t_1 + d) + (t_1 + 2d) + \cdots + t_n.$$

Deriving the Sum of an Arithmetic Series

Write the same sum twice: first in its original order and then in reverse.

$$S = t_1 + (t_1 + d) + (t_1 + 2d) + \cdots + t_n,$$
$$S = t_n + (t_n - d) + (t_n - 2d) + \cdots + t_1.$$

Add the two equations term by term. Every pair gives $t_1 + t_n$. Because there are n pairs,

$$2S = n(t_1 + t_n).$$

Dividing by 2 gives

$$\boxed{S = \frac{n}{2}(t_1 + t_n)}.$$

Since $t_n = t_1 + (n-1)d$, the same result may also be written as

$$\boxed{S = \frac{n}{2}\left[2t_1 + (n-1)d\right]}.$$

The strategy is to **change the representation**: reversing the order of the terms exposes equal pairs that were not immediately visible.

Applying the arithmetic-series formula. For the sequence $2, 5, 8, 11, 14$, we have $n = 5$, $t_1 = 2$, and $t_n = 14$. Therefore, $S = \frac{5}{2}(2 + 14) = 40$, which agrees with direct addition. The formula is a compact record of the pairing pattern revealed by writing the sum in both directions.

Then Why Do We Memorise Formulas?

There are several reasons.

1. **To keep attention on the present problem.** A learner sometimes needs a result before having the background required to prove it.
2. **To save time.** Rebuilding every result from first principles would bury the new problem beneath old work.
3. **To make implementation easier.** Computers require explicit arithmetic and logical instructions. A formula packages a relationship into a form that can often be translated directly into code.

Memorisation is useful, but it should not be mistaken for understanding. We should know what a formula means, when its assumptions apply, and where to find or reconstruct its justification, even when we do not reproduce the proof every time it is used.

3.4 Exercises

For each exercise, make the method visible as well as the answer.

1. State Pólya's four stages and explain the purpose of each.
2. In the calculation of $1 + 2 + \cdots + 100$, what made pairing the first and last terms a useful plan?
3. Which of Pólya's stages is being used when, after obtaining a proposed answer, you substitute it into the original equation to check it?
4. A room contains 12 people, and every pair of people shakes hands exactly once. Use Pólya's four stages to determine the total number of handshakes.
5. For the arithmetic sequence $5, 8, 11, \ldots, 50$, identify t_1, t_n, and d, and determine the number of terms.
6. Find the sum of the arithmetic sequence in Question 5.
7. Explain why the arithmetic-series formula is not an arbitrary rule.
8. Give two reasons why memorising an established formula may be useful even when the formula has a proof.

Optional and Non-examinable: Questions 9–10 on the Quadratic Formula

9. In the derivation of the quadratic formula, why must $a \neq 0$?
10. For $x^2 + x + 1 = 0$, identify a, b, and c, then calculate $b^2 - 4ac$. What does the result imply about real solutions?

Closing Reflection

When you use a familiar formula, what must you understand well enough to decide whether it applies?

Chapter 4

Reasoning, Evidence, and Proof

Pólya's stages help us choose and organise a solution. We now ask what kind of reasoning can establish a conclusion. This requires us to distinguish a plausible pattern from a result that follows necessarily from stated assumptions.

A proof is itself a problem-solving activity. We must understand the claim, devise an appropriate argument, carry it out clearly, and check that every assumption has been used correctly. The standard is stricter than obtaining a numerically plausible answer.

Learning Outcomes

After completing this chapter, you should be able to:

- distinguish inductive reasoning from deductive reasoning;
- distinguish mathematical intuition and evidence from proof;
- write clear and logical proofs;
- use direct proof, proof by contradiction, and mathematical induction in elementary arguments; and
- disprove a universal statement by giving a counterexample.

4.1 Inductive and Deductive Reasoning

Reasoning lets us move from information already available to a new conclusion. Two especially important forms are **inductive** and **deductive** reasoning. They do different jobs and provide different degrees of certainty.

Inductive Reasoning

Inductive reasoning uses particular observations or examples to form a broader generalisation or conjecture.

An Inductive Biological Conjecture. Suppose a biologist finds in several experiments that treated bacterial cultures grow more rapidly than comparable untreated cultures. This evidence supports the conjecture that the treatment promotes growth. It does not settle the matter: a larger experiment, another bacterial species, or an uncontrolled condition may give a different result.

Inductive reasoning is valuable for **discovering patterns and forming hypotheses**. Its conclusions vary in strength according to the quality, quantity, and representativeness of the evidence.

Deductive Reasoning

Deductive reasoning begins with one or more premises and applies valid rules of inference to reach a conclusion that must be true whenever the premises are true.

A Deductive Argument. Consider the following argument:

1. All mammals are vertebrates.
2. A dolphin is a mammal.
3. Therefore, a dolphin is a vertebrate.

If the two premises are true and the form of the argument is valid, the conclusion cannot be false. Deduction does not make its premises true; it shows what necessarily follows *if* they are accepted.

Feature	Inductive reasoning	Deductive reasoning
Typical direction	particular observations to a general claim	premises to a necessary consequence
Common purpose	discover patterns and propose hypotheses	establish what follows from assumptions
Status of conclusion	plausible or probable	certain, provided the premises are true and the reasoning is valid
Main vulnerability	limited or unrepresentative evidence	false premises or an invalid inference

The two forms of reasoning often work together. We may first notice a pattern inductively and state it as a **conjecture**. We then try to test, refine, or prove that conjecture. In an applied problem, observations may suggest a model, while deduction tells us what the model predicts.

Optional and Non-examinable: Abductive Reasoning

Abductive reasoning seeks a plausible explanation for the information available. If several leaves on a plant suddenly turn yellow, for example, one might propose nutrient deficiency as an initial explanation. This is a candidate explanation, not a certainty; competing explanations must still be investigated.

4.2 Intuition, Proof, and Certainty

Reasoning often begins with intuition. We recognise patterns, anticipate answers, and develop a sense of what ought to be true before we can fully justify it.

Mathematical Intuition

Mathematical intuition is an informed expectation about a mathematical idea or result, developed through experience with patterns, examples, diagrams, and related concepts.

Intuition helps us decide what to investigate, but it does not guarantee truth. Some correct mathematical results are surprising, while some convincing patterns eventually fail.

Mathematical Proof

A **mathematical proof** is a logically valid argument that establishes a statement from clearly stated definitions, assumptions, axioms, and results already known to be true.

Mathematical certainty is therefore conditional on the framework being used. Once the assumptions are fixed and the reasoning is valid, the conclusion follows necessarily. This is stronger than the way the word “proof” is often used in ordinary conversation.

Evidence Is Not the Same as Proof. Evidence supports a conclusion. A proof establishes that the conclusion must follow. Checking one thousand examples may provide excellent evidence for a general claim, but it does not ordinarily prove a claim about infinitely many cases. A single counterexample, on the other hand, is enough to disprove a universal claim.

4.3 Three Proof Techniques

We shall use three proof techniques:

1. A **direct proof** begins with the assumptions and uses definitions and established results to reach the required conclusion.
2. A **proof by contradiction** assumes that the claim is false and shows that this assumption leads to an impossibility.
3. A **proof by mathematical induction** establishes a first case and then proves that each case leads to the next.

Choosing a suitable technique is part of devising a plan. Writing each justified step carries out that plan, while checking the assumptions and the scope of the conclusion is part of looking back.

The examples, exercises, and assessment questions use only elementary claims. The aim is to practise the structure of each technique, not to solve difficult proof problems.

A Simple Proof by Contradiction. There is no smallest positive real number. Suppose, for contradiction, that r is the smallest positive real number. Then $r/2$ is also positive, but $r/2 < r$. This contradicts the assumption that r is the smallest. Therefore, no smallest positive real number exists.

4.4 A Recreational Problem: Is $0.999\ldots$ Less Than 1?

Consider the finite decimals $0.9 < 1$, $0.99 < 1$, $0.999 < 1$, and $0.9999 < 1$. Every decimal displayed above terminates, and each is less than 1. It is tempting to extend this pattern to $0.\overline{9} = 0.9999\ldots$, where the bar and ellipsis mean that the digit 9 repeats without end. The infinite decimal is not the same object as any finite truncation such as 0.9999.

Claim. The non-terminating decimal $0.\overline{9}$ is equal to 1.

Direct Proof. Let $x = 0.\overline{9} = 0.9999\ldots$. Multiplying both sides by 10 shifts the decimal point one place to the right, so $10x = 9.9999\ldots = 9 + x$. Therefore $10x = 9 + x$, so $9x = 9$ and hence $x = 1$. Since $x = 0.\overline{9}$, we conclude that $0.\overline{9} = 1$.

This does not say that a finite decimal such as 0.9999 equals 1. Rather, the infinitely repeating decimal and 1 are two representations of the same real number. The example shows why a pattern suggested by finite cases must be handled carefully when we pass to an infinite process.

4.5 Mathematical Patterns

Patterns often provide the starting point for mathematical discovery. We examine several cases, notice what they have in common, and use inductive reasoning to propose a general claim. We must then determine whether the pattern always holds.

Discovering a Parity Pattern. Consider the following sums:

$$\begin{aligned} 1 + 2 &= 3, & 3 + 8 &= 11, \\ 7 + 12 &= 19, & 15 + 24 &= 39. \end{aligned}$$

In every case, an odd integer is added to an even integer and the result is odd. These observations suggest the following conjecture.

Conjecture. The sum of an odd integer and an even integer is always odd.

The examples provide evidence, but they cannot cover every possible pair of integers. To prove the conjecture, we use the definitions of odd and even integers.

Even and Odd Integers

An integer is **even** if it can be written as $2n$ for some integer n. An integer is **odd** if it can be written as $2m + 1$ for some integer m.

Direct Proof. Let $2m + 1$ be any odd integer and let $2n$ be any even integer, where m and n are integers. Their sum is $(2m + 1) + 2n = 2m + 2n + 1 = 2(m + n) + 1$. Because the sum of two integers is an integer, $m + n$ is an integer. The expression $2(m + n) + 1$ therefore has the form $2k + 1$ for an integer $k = m + n$. By definition, it is odd. Hence, the sum of any odd integer and any even integer is odd.

The examples helped us discover the claim; the proof establishes it for all integers, including cases not tested. A common mathematical sequence is examples $\longrightarrow$ pattern $\longrightarrow$ conjecture $\longrightarrow$ proof.

4.6 Mathematical Induction

Inductive reasoning uses particular cases to suggest a general claim. **Mathematical induction** is different: it is a method for proving a claim about every natural number. The method has two parts.

1. **Base case.** Show that the claim is true for the first value.
2. **Inductive step.** Assume that the claim is true for an arbitrary natural number k, and prove that it is true for $k + 1$.

If both parts are established, the claim holds for every natural number from the starting value onwards. The logic is similar to a row of dominoes: the base case knocks over the first domino, and the inductive step shows that each fallen domino knocks over the next one.

Example. For every positive integer n,

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2.$$

Proof by Mathematical Induction. For $n = 1$, the left-hand side is 1, and the right-hand side is $1^2 = 1$. The claim is therefore true for the base case.

Now assume that the claim is true for some positive integer k; that is, assume

$$1 + 3 + 5 + \cdots + (2k - 1) = k^2.$$

For $k + 1$, the next odd number is $2(k + 1) - 1 = 2k + 1$. Adding it to both sides of the assumed statement gives

$$1 + 3 + 5 + \cdots + (2k - 1) + (2k + 1) = k^2 + 2k + 1 = (k + 1)^2.$$

Thus, if the claim is true for k, it is true for $k + 1$. Since the base case is true, the identity holds for every positive integer n.

4.7 Disproof by Counterexample

Consider the claim that every prime number is odd. The primes 3, 5, 7, 11, and 13 support it, but 2 is both prime and even. Thus, 2 is a **counterexample**: it satisfies the condition of the claim but not its conclusion. One counterexample disproves a universal statement; confirming examples alone do not prove it.

4.8 Exercises

First distinguish the forms of reasoning, then construct short arguments of your own.

1. Explain why an inductive conclusion can be reasonable without being certain.
2. In the dolphin argument, identify the two premises and the conclusion. Is the reasoning inductive or deductive?

3. Give a biological example in which observations might support a hypothesis without establishing it with certainty.
4. Explain the difference between evidence for a universal claim and a proof of that claim.
5. Use proof by contradiction to show that there is no largest integer.
6. In the algebraic proof of $0.\overline{9} = 1$, explain why $9.9999\ldots = 9 + 0.9999\ldots$.
7. A student calculates $2 + 3 = 5$, $4 + 7 = 11$, and $10 + 9 = 19$. What general pattern might the student conjecture? Why do the three examples not prove the claim?
8. In the proof about parity, explain why writing the sum as $2(m + n) + 1$ completes the argument.
9. Prove directly that the sum of two odd integers is even.
10. Use mathematical induction to prove that, for every positive integer n,

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

11. Give a counterexample to the claim that every odd integer is prime. Explain why one example is sufficient.

Closing Reflection

When a mathematical claim seems convincing, what would you still need to check before calling your argument a proof?

Chapter 5

Managing and Describing Data

Biology produces data at almost every scale, but a page of measurements is not yet an argument. The measurements have to be gathered, organised, checked, displayed, and summarised before they can support a conclusion. Here we stop at data management and description; probability and regression come in Chapter 6.

Learning Outcomes

After completing this chapter, you should be able to:

- distinguish observations, variables, populations, and samples;
- organise a dataset and document how its values are defined;
- recognise common sources of bias, error, and missingness;
- select and interpret appropriate tables and graphs, by hand or with available software;
- calculate and interpret measures of centre and dispersion; and
- interpret quartiles, the IQR, and box plots.

The working sequence is

$$\text{question} \longrightarrow \text{collection} \longrightarrow \text{management} \longrightarrow \text{description}.$$

No calculation can correct a poorly framed question or badly collected data.

5.1 Understanding and Managing Data

5.1.1 From a Question to a Dataset

A germination and growth dataset. Suppose we want to investigate whether daily light exposure is associated with germination and, for seeds that germinate, seedling height after 14 days. We might record data in the following form.

Seed ID	Light per day (hours)	Height after 14 days (cm)	Leaf colour	Germinated?
S01	4	7.1	pale green	yes
S02	6	9.3	green	yes

Seed ID	Light per day (hours)	Height after 14 days (cm)	Leaf colour	Germinated?
S03	8	11.0	dark green	yes
S04	8	—	not applicable	no

Each row describes one planted seed, and each column records one feature of it.

Observations and Variables

An **observation**, sometimes called a case or record, is one unit about which data have been collected. In the table, each planted seed is an observation.
A **variable** is a characteristic recorded for each observation. Light, height, leaf colour, and germination status are variables.

The seed ID distinguishes one observation from another; it is an **identifier** rather than a biological measurement. For S04, height and leaf colour are unavailable because no seedling developed. The reason should be recorded explicitly: height was not measured, and there was no leaf whose colour could be classified. Neither entry is a measured zero.

5.1.2 Types of Variables

The type of a variable affects which summaries and graphs are meaningful.

Variable type	Meaning	Biological example
categorical, nominal	labels without a natural order	blood type, species, habitat
categorical, ordinal	categories with a meaningful order	mild, moderate, severe
numerical, discrete	counts, usually whole numbers	number of colonies, number of leaves
numerical, continuous	measurements on a scale	mass, temperature, concentration

Numbers do not automatically make a variable numerical in the statistical sense. Codes such as `1 = control` and `2 = treatment` still represent categories. Averaging these codes would have no sensible scientific meaning.

A Variable Is Defined by Meaning, Not Appearance

A phone number, student number, or specimen code consists of digits, but it is an identifier. Conversely, categories such as “mild”, “moderate”, and “severe” contain no digits but have a meaningful order.

5.1.3 Populations and Samples

Populations and Samples

The **population** is the complete collection of observations about which we want to draw a conclusion.
A **sample** is the subset of the population that we actually observe.

If our question concerns all first-year BS Biology students at BulSU, those students form the population of interest. The 40 students who answer a survey form a sample. A **census** attempts to observe every member of the population, although incomplete responses may prevent it from being truly complete.

We use a sample because measuring an entire population may be too expensive, slow, destructive, or impossible. The central difficulty is that a sample must represent the population well enough for the intended conclusion.

5.1.4 Gathering Data

Data may arise from the following sources:

- **experiments**, in which researchers deliberately impose conditions or treatments;
- **observational studies**, in which researchers record existing conditions without assigning treatments;
- **surveys and interviews**;
- **administrative records**;
- **sensors, instruments, and images**; or
- **published databases** collected by another organisation.

Each source creates different limitations. An experiment with suitable controls can support a causal conclusion more readily than an observational study. A survey can reach many people but may suffer from non-response or misreporting. A sensor may be precise but poorly calibrated.

Bias

Bias is a systematic tendency for a method to produce results that differ from the quantity or relationship we intend to study.

Common sources include:

- **selection bias:** the sampled observations systematically differ from the target population;
- **non-response bias:** those who do not respond differ meaningfully from those who do;
- **measurement bias:** an instrument or procedure consistently records values too high, too low, or in an otherwise distorted way;
- **response bias:** participants alter or misreport their answers; and
- **confounding:** another variable influences both variables being studied, creating or obscuring an apparent relationship.

Increasing the sample size does not automatically remove bias. A very large biased sample can provide a very precise answer to the wrong question.

5.1.5 Organising and Documenting Data

A useful working convention is:

- one row for each observation;
- one column for each variable; and
- one value in each cell.

This arrangement makes sorting, filtering, graphing, and analysis more reliable. It also helps other people understand what one row means.

Before analysis, check for the following:

- duplicated observations;
- impossible values, such as a negative body mass;
- inconsistent units, such as mixing grams and kilograms;
- inconsistent category labels, such as `Female`, `female`, and `F`;
- typing or transcription errors;
- missing values; and
- measurements recorded with inappropriate precision.

Create a data dictionary. A **data dictionary** records the meaning of each variable, its units, allowed values, category codes, and the meaning of missing-value symbols.

Variable	Meaning	Unit or allowed values
`height14`	height 14 days after planting	centimetres
`light`	average light exposure per day	hours
`germinated`	whether germination occurred	yes/no

Good documentation is part of the mathematics because an ambiguous variable cannot be interpreted precisely.

5.1.6 Why a Value Is Missing Matters

A missing value may have several causes:

- a participant declined to answer;
- an instrument failed;
- the measurement was not applicable;
- an observation was lost;
- the value was never collected; or
- the organism did not survive long enough to be measured.

These situations are scientifically different. Replacing every missing value with zero changes the data and may create a false conclusion. Deleting all incomplete observations can also introduce bias if the missingness is related to the phenomenon being studied.

Record missing values consistently, find out why they are missing, and state how you handled them.

5.1.7 Ethical Data Management

Data about people may contain private or sensitive information. Depending on the study, this calls for informed consent, restricted access, removal of direct identifiers, secure storage, and restraint in what is collected.

The Philippine Data Privacy Act

In the Philippines, Republic Act No. 10173 governs the processing of personal data. It is the *Data Privacy Act of 2012*. Data processing must follow three general principles: **transparency**, **legitimate purpose**, and **proportionality**. Health information is sensitive personal information and therefore requires particular care.

Collect only the information required for the stated purpose, explain how it will be used, restrict access, store it securely, and retain it only for as long as necessary. Instructional datasets should be anonymised or synthetic unless the use of identifiable information has been expressly authorised.

Never fabricate an observation, silently change an inconvenient value, or remove an outlier merely because it weakens the conclusion you hoped to reach. Cleaning data means correcting documented errors and recording each decision. It does not mean making the dataset tell a preferred story.

Scientific work depends on an honest account of what was observed. Fabricated or manipulated data can misdirect later research, waste resources, and, in medical settings, expose people to harm.

Digital Checks and Human Judgment. Software may flag a duplicated or altered image, but the flag is only a reason to inspect the original files and laboratory record. It does not establish misconduct on its own.

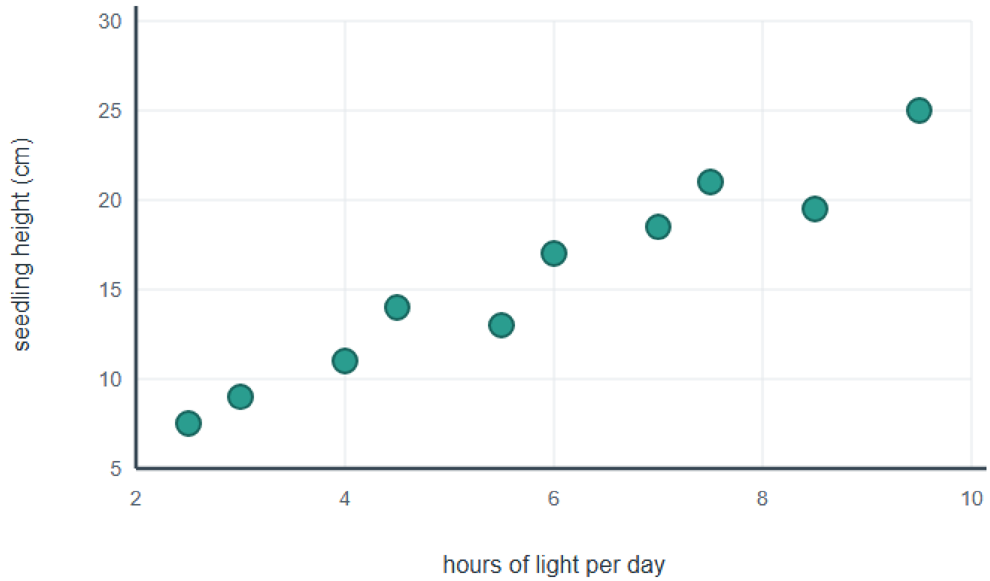
5.2 Representing and Describing Data

5.2.1 Tables, Graphs, and Charts

A good display should answer a question. Decoration is secondary.

Purpose	Usually suitable display
compare counts or proportions across categories	bar chart
show the distribution of one numerical variable	histogram or box plot
show change over time	line graph
investigate two numerical variables	scatterplot
display exact small counts	frequency table

Light exposure and seedling height. A scatterplot represents each observation by one point. Its horizontal position records one numerical variable and its vertical position records another. In the example below, each point represents one seedling.



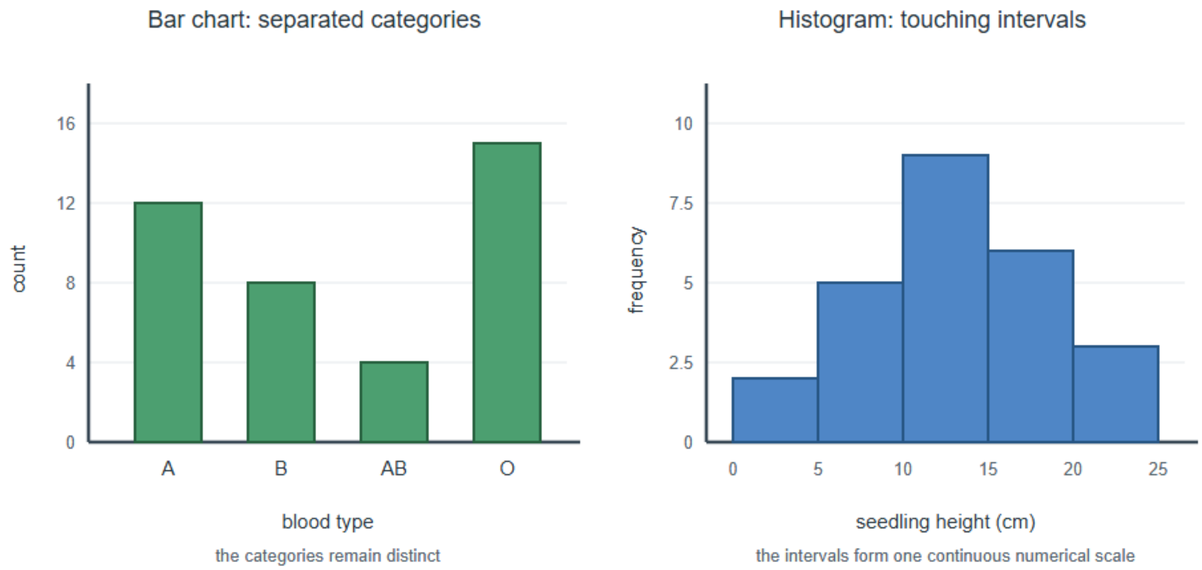
An illustrative scatterplot of daily light exposure and seedling height. The points suggest that seedlings receiving more light tend to be taller, but the relationship is not exact.

Tools You May Encounter. You will probably use software to manage data and produce statistical figures in later study or employment. Common tools include **R**, often used through the **RStudio** interface; **Python**, with libraries such as **pandas** for data management and **Matplotlib** or **seaborn** for graphics; and **MATLAB**. You may also encounter packages with graphical interfaces, such as **SPSS**, **jamovi**, or **GraphPad Prism**.

I do not assume any statistical programming in this course. You should, however, recognise these names and know what the tools are for. Software can calculate and draw; it cannot decide whether a graph answers the scientific question or whether an interpretation is justified.

A **bar chart** keeps categories separate. A **histogram** groups numerical values into adjacent intervals, so its bars touch. Treating a histogram as a bar chart—or a bar chart as a histogram—confuses two different kinds of data.

Bar chart and histogram.



A bar chart and a histogram answer different questions. The gaps in the bar chart separate blood-type categories; the touching histogram bars represent adjacent intervals on a continuous height scale.

When reading any display, ask the following questions:

1. What population and variables are represented?
2. Are the axes labelled and are units given?
3. Does the vertical axis begin at zero, and if not, does the truncation exaggerate a difference?
4. Are intervals equally wide?
5. Are counts, percentages, or rates being shown?
6. Is any group or time period missing?

Beginning a vertical axis at zero is particularly important for a bar chart, because bar length represents magnitude. A scatterplot or line graph need not always begin at zero, but any restricted scale should be clearly labelled and chosen for a stated reason.

A Graph Can Be Numerically Correct and Still Misleading

Changing an axis scale, choosing unusual interval widths, omitting a category, or displaying totals instead of rates can create a distorted impression without changing the underlying numbers.

5.2.2 Measures of Central Tendency

A measure of central tendency describes a typical or central location of a dataset. Different measures answer different questions.

Measures of Centre

Mean. For numerical observations $x_1, x_2, \dots, x_n$, the **arithmetic mean** is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Median. For ordered numerical observations, the **median** is the middle value when the number of observations is odd. When it is even, the median is the arithmetic mean of the two central values.
Mode. The **mode** is the value or category occurring most frequently. A dataset may have one mode, several modes, or no unique mode.

Germination times. Consider the synthetic germination times $4, 5, 5, 6, 10$, measured in days. Their mean is $\bar{x} = (4 + 5 + 5 + 6 + 10)/5 = 6$, their median is 5, and their mode is 5. The value 10 raises the mean, whereas the median remains 5 and is less sensitive to this unusually late germination time.

The mean uses the magnitude of every observation and is useful when totals matter. The median is resistant to unusually large or small values and is often more representative for a skewed distribution. The mode is particularly useful for categorical data, for which a numerical mean may not exist.

5.2.3 Weighted Mean

Some values represent groups of different sizes or otherwise carry different weights. If values $x_1, \dots, x_k$ have nonnegative weights $w_1, \dots, w_k$, with $\sum_{i=1}^{k} w_i > 0$, their weighted mean is

$$\bar{x}_w = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i}.$$

Combined germination percentage. Suppose 80% of 20 seeds germinate in Tray A and 60% of 30 seeds germinate in Tray B. The combined germination percentage is not the unweighted average $(80 + 60)/2 = 70\%$ unless both trays contain the same number of seeds. Weighting by the tray sizes gives $\frac{20(80)+30(60)}{20+30} = 68\%$.

Equivalently, 16 seeds germinated in Tray A and 18 in Tray B, giving $34/50 = 68\%$ overall.

5.2.4 Measures of Dispersion

Two datasets can have the same mean but very different variability. **Dispersion** describes how widely the observations are spread. Variance and standard deviation measure spread about the mean, while the range and interquartile range describe it in other ways.

Range

The **range** is maximum − minimum.
It is easy to calculate but depends only on the two extreme observations.

Calculating the range. The germination times $4, 5, 5, 6, 10$ have range $10 - 4 = 6$ days.

Variance and standard deviation use the distance of every observation from the mean.

Variance and Standard Deviation

Population. For a complete population of size N with mean μ,

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2, \qquad \sigma = \sqrt{\sigma^2}.$$

Sample. For a sample of size $n > 1$ with mean $\bar{x}$,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad s = \sqrt{s^2}.$$

The deviations are squared so that values below and above the mean do not cancel. The standard deviation returns the result to the original unit. If height is measured in centimetres, variance is measured in square centimetres, while standard deviation is again measured in centimetres.

Sample variance and standard deviation. For the sample $4, 5, 5, 6, 10$, the mean is 6 and the squared deviations sum to

$$(4-6)^2 + (5-6)^2 + (5-6)^2 + (6-6)^2 + (10-6)^2 = 22.$$

Therefore,

$$s^2 = \frac{22}{5-1} = 5.5 \qquad \text{and} \qquad s = \sqrt{5.5} \approx 2.35 \text{ days.}$$

Do Not Report a Statistic Without Its Context

"The standard deviation is 2.35" is incomplete. We need to know the variable, unit, population or sample, and circumstances of collection. Here it is a sample standard deviation of approximately 2.35 **days** for the five synthetic germination times.

5.3 Relative Position: Quartiles and Box Plots

5.3.1 Percentiles and Quartiles

A **percentile** describes relative position in an ordered dataset. Informally, the pth percentile is a value at or below which approximately $p\%$ of the observations lie.

The three quartiles divide ordered data into four parts:

- Q_1: first quartile, near the 25th percentile;
- Q_2: second quartile, equal to the median; and
- Q_3: third quartile, near the 75th percentile.

The **interquartile range** is $\mathrm{IQR} = Q_3 - Q_1$.

It describes the spread of the middle half of the data and is less sensitive to extreme observations than the ordinary range.

Quartile Conventions Differ

Software packages and textbooks use several rules for locating quartiles and percentiles, especially in small datasets. Always state or identify the convention being used when an exact value matters.

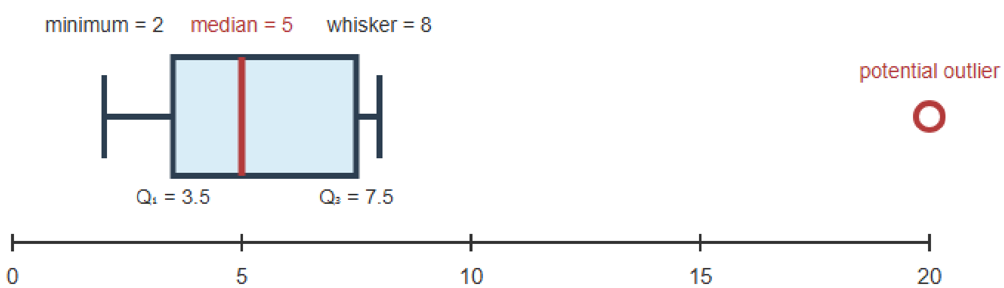
5.3.2 Box-and-Whisker Plots

A box-and-whisker plot, or **box plot**, summarises a distribution using Q_1, the median, and Q_3. Under the common 1.5 IQR rule, the whiskers extend to the smallest and largest observations that are not flagged as potential outliers; flagged values are shown separately.

Identifying a potential outlier. Consider 2, 3, 4, 5, 5, 6, 7, 8, 20.

Using the convention that excludes the median when splitting an odd-sized dataset, $Q_1 = 3.5$, $Q_2 = 5$, and $Q_3 = 7.5$, so IQR $= 4$. A common exploratory rule flags values below $Q_1 - 1.5(\text{IQR}) = -2.5$ or above $Q_3 + 1.5(\text{IQR}) = 13.5$.

The value 20 is therefore flagged as a potential outlier.



Box plot for the dataset. The whiskers extend from 2 to 8; the box runs from Q1 = 3.5 to Q3 = 7.5, with median 5. The value 20 is shown separately as a potential outlier.

An Outlier Is a Question, Not an Automatic Deletion

The value may be a typing error, an instrument failure, or a genuine and scientifically important observation. Investigate it. Do not remove it merely because a rule placed it beyond a fence.

If these were marks on a test, an extreme value could reflect unusual performance, illness, misunderstood instructions, or a recording error. The box plot cannot tell us which explanation is correct. It only identifies values that merit checking.

5.4 Exercises

Work with the data as well as the formulas: define what was recorded, account for what is missing, and explain each conclusion.

1. In the germination and growth table, identify the observations, numerical variables, categorical variables, identifier, and entries that are unavailable for S04.

2. Explain why coding `control = 1` and `treatment = 2` does not make treatment group a numerical variable.
3. Distinguish a population from a sample using a biological example.
4. Explain why increasing the size of a biased sample does not necessarily produce a trustworthy conclusion.
5. For the values $4, 5, 5, 6, 10$, calculate the mean, median, mode, range, sample variance, and sample standard deviation. Which measure of centre is less affected by the value 10?
6. Two trays contain 10 and 40 seeds. Their germination percentages are 90% and 60%, respectively. Find the combined germination percentage.
7. Why should an observation flagged by the 1.5 IQR rule be investigated rather than automatically deleted?
8. Explain why the bars of a histogram ordinarily touch while the bars of a bar chart ordinarily do not.
9. Examine the seedling scatterplot above. Describe its overall pattern and identify the two variables and their units.
10. A dataset contains the columns `specimen_id`, `species`, and `mass_g`. Write a brief data-dictionary entry for each column, including its meaning and its unit or allowed values.
11. A planted seed did not germinate, so no seedling height was measured after 14 days. Explain why recording its height as zero could be misleading. How should the value and the reason for its absence be documented?
12. A student project proposes collecting participants' names and medical histories even though its question requires only age groups without direct identifiers. Identify two changes that would make the proposed data collection more proportionate and protective of privacy.

Closing Reflection

Think of a dataset you have encountered. What could one missing value mean, and how might that meaning affect an analysis?

Chapter 6

Probability, Association, and Prediction

Chapter 5 ended with descriptions of observed data. We now ask where an observation sits within a distribution, how two variables move together, and what a fitted model might predict. Every answer depends on assumptions; sound arithmetic cannot rescue an unsuitable model.

Learning Outcomes

After completing this chapter, you should be able to:

- calculate and interpret z-scores;
- interpret elementary probabilities under a normal model;
- describe a scatterplot and interpret a linear correlation coefficient;
- use a least-squares line to make a prediction;
- distinguish association from causation and interpolation from extrapolation; and
- state what statistical evidence can support when explaining or defending a decision.

6.1 Standardised Position and Probability

6.1.1 z-Scores

A z-score expresses how far an observation lies from the mean in standard-deviation units. For a population with $\sigma > 0$, $z = (x - \mu)/\sigma$; when sample statistics with $s > 0$ are used, $z = (x - \bar{x})/s$.

Interpreting a z-score. If a plant has height 72 cm in a population with mean 60 cm and standard deviation 6 cm, then $z = (72 - 60)/6 = 2$. The plant is two standard deviations above the mean. A z-score gives relative position, not probability. Converting it to a probability requires a model for the distribution.

6.1.2 Elementary Probability

For an event A, $0 \le P(A) \le 1$.

A probability near 0 describes an unlikely event under the stated model; a probability near 1 describes a likely one. The complement rule is $P(A^c) = 1 - P(A)$.

Probabilities may be expressed as proportions, decimals, or percentages. Every probability statement should identify the event and the conditions under which it is being calculated.

Probability Is Conditional on a Model

A probability is not a promise about one individual. It describes uncertainty under specified assumptions or a long-run pattern under repeated comparable conditions.

6.1.3 The Normal Distribution

A normal distribution is a symmetric, bell-shaped probability model determined by a mean μ and a standard deviation $\sigma > 0$. Under a normal model:

- approximately 68% of values lie within one standard deviation of the mean;
- approximately 95% lie within two standard deviations; and
- approximately 99.7% lie within three standard deviations.

These percentages are useful approximations, not universal laws for every dataset. Before using a normal model, we should examine whether the distribution is reasonably symmetric and whether strong outliers or bounds make the model implausible.

Interpreting a Large z-Score

A large value of $|z|$ means that an observation lies far from the reference mean relative to the stated standard deviation. If the distributional model and reference values are appropriate and the measurements are reliable, such an observation is unusual—but unusual observations can still occur.

One large z-score does not establish that the reference mean is wrong. It might represent a genuine rare observation, a measurement or recording error, an unsuitable distributional model, or incorrect reference values. If many independent, well-measured observations lie unusually far from the reference mean in the same direction, that pattern may give us reason to question the assumed mean or the wider model. Formally evaluating that question requires sampling distributions and hypothesis testing, which lie beyond the scope of this course.

Optional and Non-examinable: Why Normal Distributions Appear So Often

Biological measurements themselves need not follow a normal distribution. However, under suitable conditions, the distribution of the **sample mean** becomes approximately normal as the sample size increases. This result is called the **central limit theorem**.

The theorem helps explain why normal models are widely used when estimating population means. It does not correct selection bias or poor measurement; increasing the sample size can leave those problems intact. Nor is there a universal sample size at which the approximation suddenly becomes valid. The required size depends on the original distribution and the way the observations were collected.

The precise conditions, standard errors, and inferential applications of the theorem belong to a later statistics course.

6.1.4 Worked Normal-Model Example

Calories burned under a normal model. Assume that calories burned during a one-hour brisk walk at 6.4 km/h follow a normal distribution with mean 300 calories and standard deviation 8 calories. Let $X \sim N(300, 8^2)$.

This notation means that X is being modelled as normal with mean 300 and variance 8^2.

More Than 280 Calories. Standardising gives $z = (280 - 300)/8 = -2.5$. Using a normal table or statistical software, $P(X > 280) = P(Z > -2.5) \approx 0.9938$.

Under the model, about 99.4% of comparable one-hour walks would burn more than 280 calories.

Less Than 293 Calories. Here, $z = (293 - 300)/8 = -0.875$, so $P(X < 293) \approx 0.1908$.

Under the model, about 19.1% would burn fewer than 293 calories.

Between 285 and 320 Calories. The corresponding z-scores are $(285 - 300)/8 = -1.875$ and $(320 - 300)/8 = 2.5$. Therefore, $P(285 < X < 320) \approx 0.9634$.

About 96.3% would fall in that interval under the assumed normal model.

The Arithmetic Can Be Correct While the Model Is Wrong

These answers depend on the assumptions that the stated mean and standard deviation are appropriate for the population and that a normal model is reasonable. Real calorie expenditure also depends on body mass, physiology, terrain, measurement method, and other conditions. Statistical interpretation must carry the assumptions along with the answer.

6.2 Relationships, Regression, and Prediction

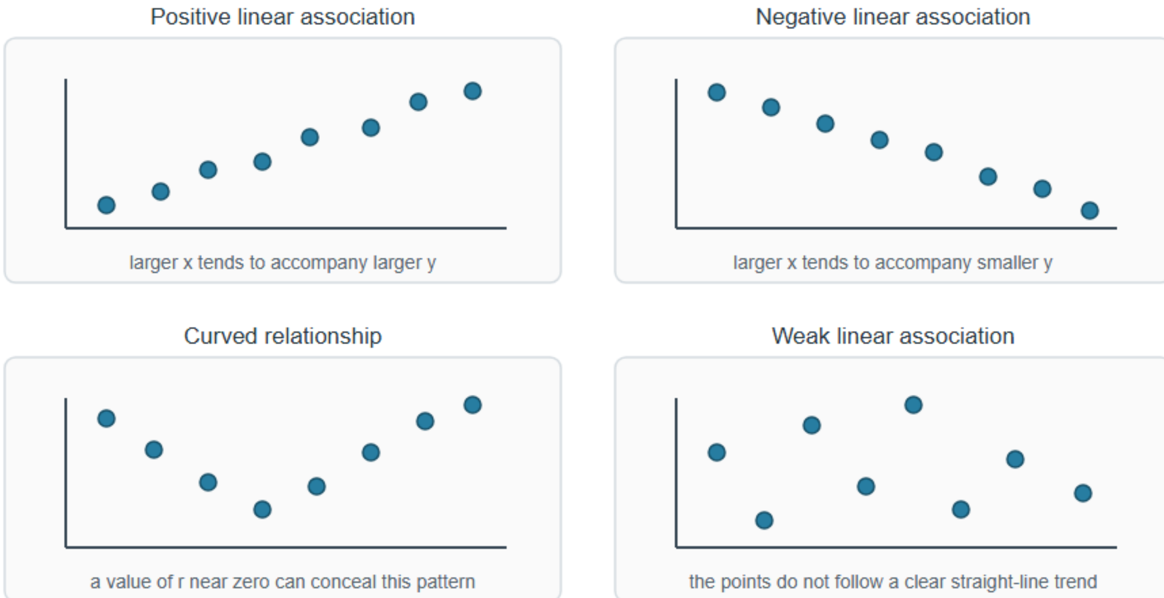
6.2.1 Begin With a Scatterplot

When two numerical variables are recorded on the same observations, a scatterplot places one variable on the horizontal axis and the other on the vertical axis. We look for:

- **direction:** positive, negative, or neither;
- **form:** linear, curved, clustered, or another pattern;
- **strength:** how tightly the points follow that form; and
- **unusual observations:** outliers or influential points.

A correlation coefficient should not be interpreted without first looking at the scatterplot. A single number can hide curvature, clusters, data-entry errors, or an observation exerting excessive influence.

Four scatterplot patterns.



Four patterns that a scatterplot may reveal. Direction and strength must be judged with the form of the relationship in view; a curved pattern can be strong even when its linear correlation is small.

6.2.2 Linear Correlation

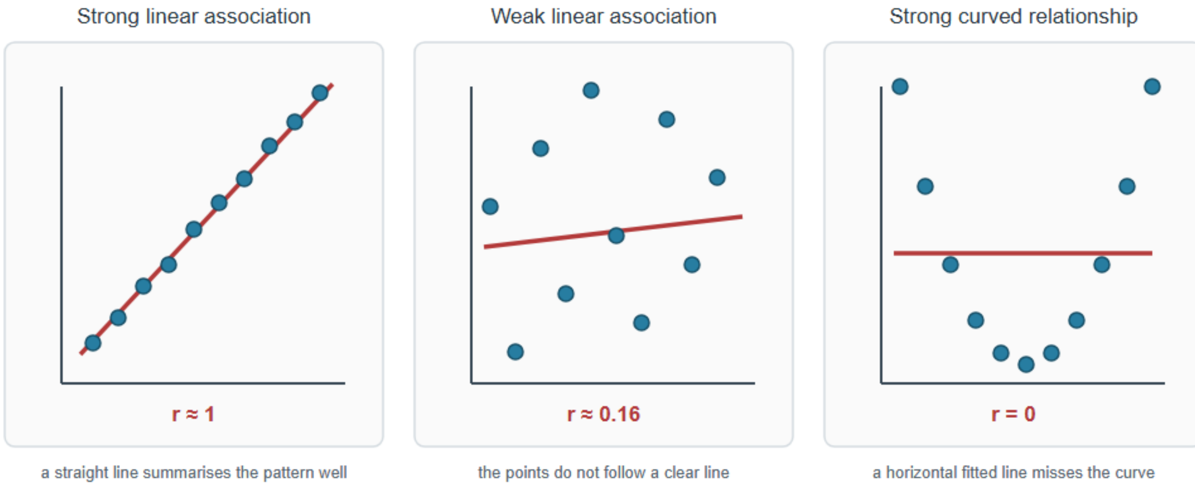
The Pearson linear correlation coefficient is

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

This coefficient is defined provided that neither variable is constant; under that condition, $-1 \le r \le 1$.

- A positive r indicates that larger values of one variable tend to occur with larger values of the other.
- A negative r indicates that larger values of one tend to occur with smaller values of the other.
- A value near zero indicates weak **linear** association, although a strong curved relationship may still exist.

What Pearson's r can miss.



Pearson's correlation describes linear association. The third dataset has a strong quadratic pattern, but its upward and downward tendencies cancel in the linear coefficient.

The third panel shows why a correlation near zero does not establish that two variables are unrelated. The small coefficient reflects the mismatch between a straight line and the curved data, not an absence of structure.

Correlation has no physical unit and is unchanged by converting, for example, centimetres to metres. It measures linear association, not the size of a causal effect.

Correlation Does Not Imply Causation

An association may arise because X affects Y, because Y affects X, because another variable affects both, because of selection or measurement bias, or because of chance. Establishing causation requires study design and subject-matter reasoning, not merely a large value of r.

6.2.3 Least-Squares Regression

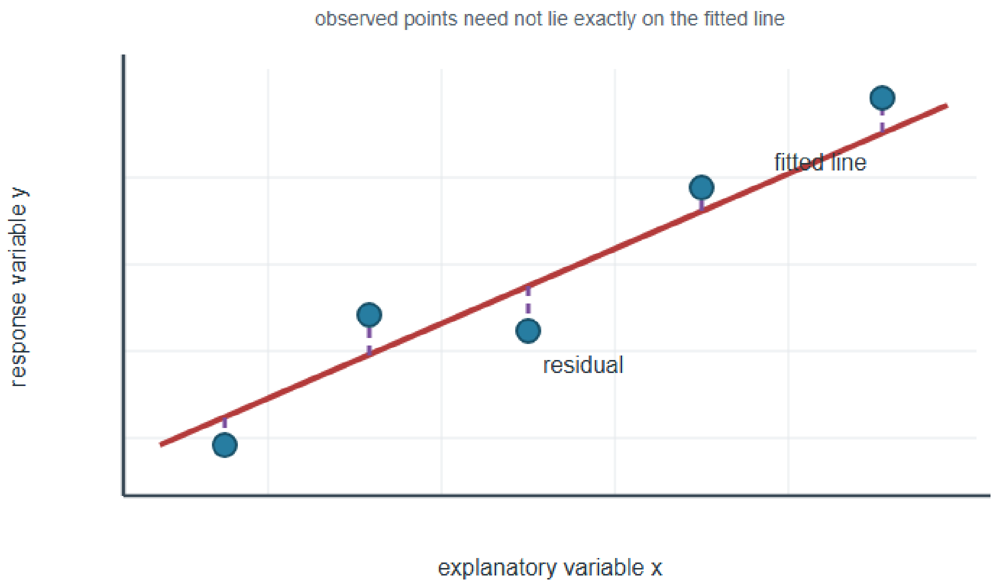
A simple linear regression line predicts a response variable y from an explanatory variable x:

$$\hat{y} = b_0 + b_1 x.$$

The symbol $\hat{y}$, read "y-hat", is the predicted value. The slope b_1 is the predicted change in y associated with a one-unit increase in x. The intercept b_0 is the predicted value when $x = 0$, although that interpretation may be meaningless if zero lies far outside the observed range.

The **residual** for observation i is $e_i = y_i - \hat{y}_i$.

Residuals from a fitted line.



A residual is the vertical difference between an observed response and the response predicted by the fitted line.

The least-squares line chooses b_0 and b_1 to minimise $\sum_{i=1}^{n} e_i^2$.

Squaring prevents positive and negative residuals from cancelling and imposes a strong penalty on large errors.

When $s_x > 0$ and $s_y > 0$, the simple linear regression coefficients may be written as

$$b_1 = r\frac{s_y}{s_x} \qquad \text{and} \qquad b_0 = \bar{y} - b_1\bar{x}.$$

If $s_x > 0$ but every observed y-value is equal, then $s_y = 0$, the fitted line is horizontal, and Pearson's r is undefined.

Optional and Non-examinable: Where $\mathbf{b_0}$ and $\mathbf{b_1}$ Come From

To derive the formulas, treat the sum of squared residuals as a function of the two unknown coefficients and use multivariable calculus:

$$S(b_0, b_1) = \sum_{i=1}^{n}(y_i - b_0 - b_1x_i)^2.$$

At the least-squares solution, the partial derivatives with respect to b_0 and b_1 are zero:

$$\frac{\partial S}{\partial b_0} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1x_i) = 0,$$

$$\frac{\partial S}{\partial b_1} = -2\sum_{i=1}^{n}x_i(y_i - b_0 - b_1x_i) = 0.$$

Solving these two equations gives

$$b_0 = \bar{y} - b_1\bar{x}$$

and

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

When $s_y > 0$, this is also $b_1 = r(s_y/s_x)$.
Because a sum of squares is a convex function of the coefficients, this stationary point is the minimum; it is unique provided that the observed x-values are not all identical. A complete treatment belongs to a course in multivariable calculus, linear algebra, or mathematical statistics.

Software will ordinarily perform the arithmetic. You still have to identify the variables, inspect the graph, interpret the coefficients and residuals, judge the model, and state its limitations.

6.2.4 Interpolation and Extrapolation

Interpolation predicts within the range of observed explanatory values. **Extrapolation** predicts beyond that range. Extrapolation is more dangerous because the relationship may change where no data were collected.

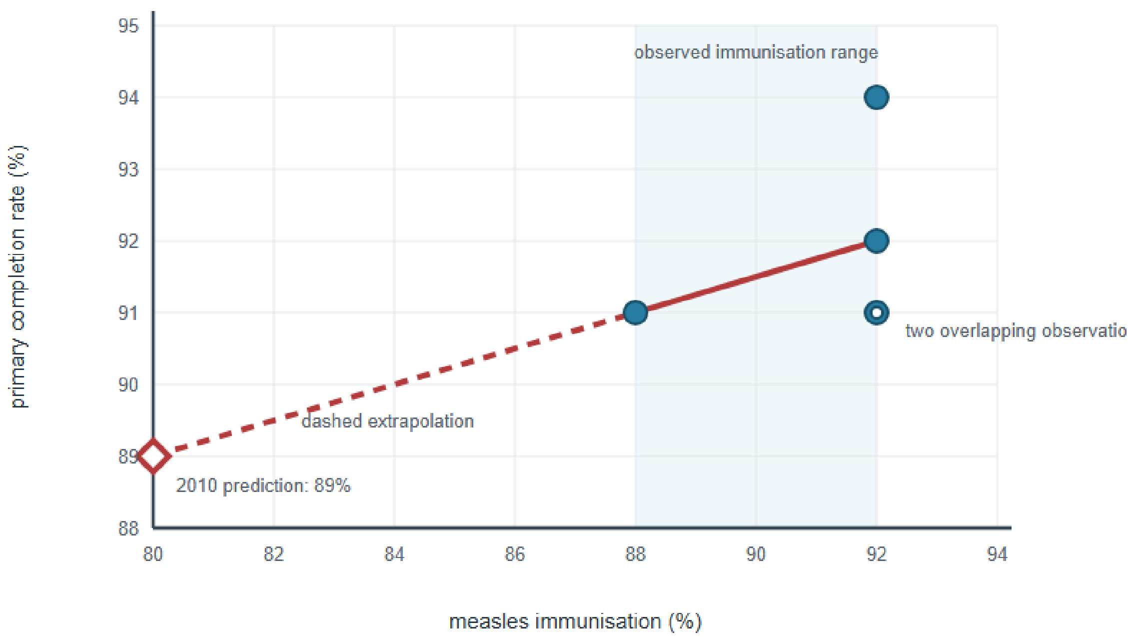
A regression equation is not a law of nature. A line that describes seedlings receiving four to eight hours of light should not automatically be used to predict growth under zero or twenty-four hours of light.

6.2.5 Illustrative Example: Immunisation and School Completion

Immunisation and school completion. The table below gives an illustrative dataset. The primary completion rate for 2010 is missing.

Year	Measles immunisation (% of children aged 12–23 months)	Primary completion rate (%)
2005	92	94
2006	92	91
2007	92	91
2008	92	92
2009	88	91
2010	80	missing

Using the five complete rows, the least-squares line predicting completion rate y from immunisation percentage x is approximately $\hat{y} = 69 + 0.25x$. Substituting $x = 80$ gives $\hat{y} = 69 + 0.25(80) = 89$.



The fitted line is supported only over the narrow observed range from 88% to 92% immunisation. Predicting at 80% extends the line well beyond the observed data.

The fitted equation gives 89%, but that number should not be treated as a reliable forecast. The correlation in the five complete rows is only about $r = 0.34$, the fitted dataset contains only five complete observations, most explanatory values are identical, and $x = 80$ lies outside the observed range from 88 to 92. The calculation is an extrapolation.

Nor does the regression show that increasing immunisation causes higher school completion. Both variables may be influenced by funding, access to services, geography, policy, and many other conditions.

Reporting the prediction. A careful report might say:

> The fitted line gives a numerical prediction of approximately 89%, but the estimate is highly uncertain because it extrapolates from only five complete observations with a weak linear association. It should not be used for an important decision without more suitable data.

6.3 Using Statistical Evidence in Decisions

A statistical result addresses only the quantities included in the analysis. Costs, feasibility, ethics, uncertainty, and the consequences of error must be considered separately.

When presenting statistical evidence:

1. state the question and population clearly;
2. explain how the data were gathered;

3. identify the variables and units;
4. use appropriate displays and summaries;
5. separate observed facts from model-based predictions;
6. disclose missing data and analytical choices;
7. discuss bias, uncertainty, and limitations; and
8. avoid causal language unless the design justifies it.

Give readers enough information to inspect the source, method, and limitations of each numerical claim.

6.4 Quantitative Study Proposal

Planning a university shuttle study. Suppose the university is considering free shuttle services from selected points outside the campus to reduce congestion. The administration asks for evidence before deciding whether and how to proceed.

A proposal should describe the following:

- the precise decision or research question;
- the target population;
- the sampling method and intended sample size;
- the variables to be collected;
- the survey, observation, or measurement procedure;
- how data will be coded, stored, checked, and documented;
- the graphs and statistical methods that may be appropriate;
- privacy and consent safeguards;
- likely sources of bias and uncertainty; and
- the form in which results will be communicated.

A proposal describes what will be done. It contains no invented results and makes no promise that the evidence will favour the shuttle service.

6.5 Working With Technology

Calculators, spreadsheets, and statistical software can:

- sort and filter observations;
- calculate summary statistics;
- produce graphs;
- evaluate normal probabilities;
- calculate correlation coefficients; and
- fit regression models.

Technology saves time and reduces routine arithmetic, but it does not decide whether a variable was measured well, whether a graph is misleading, whether missing data create bias, or whether a causal conclusion is justified.

A Software Output Is Not an Interpretation

Always translate output into a complete statement about the variables, population, units, model, and limitations. Copying a table of numbers from software does not amount to statistical reasoning.

6.6 Exercises

For each calculation or classification, state the limits of the conclusion.

1. A measurement has value 45, while its population has mean 39 and standard deviation 3. Calculate and interpret its z-score.
2. Suppose $X \sim N(100, 15^2)$. Use a normal table or statistical software to calculate $P(X > 130)$ and interpret the result under the model.
3. Under a stated normal model, a probability calculation is correct. Give two reasons why its practical interpretation might nevertheless be poor.
4. Explain why a correlation near zero does not prove that two variables are unrelated.
5. Distinguish interpolation from extrapolation. Which was used in the illustrative immunisation example, and why?
6. A fitted line is $\hat{y} = 4 + 2.5x$. Predict y when $x = 6$ and interpret the slope in context if x is hours of light and y is seedling height in centimetres.
7. A regression between hours of sleep and examination score produces a positive correlation. Give at least two reasons this does not prove that increasing sleep will cause every student's score to rise.
8. For the shuttle proposal, name one variable, one possible source of bias, and one ethical or privacy consideration.

Closing Reflection

Before trusting a statistical prediction, which assumption about the data or model would you check first?

Chapter 7

Interest, Credit, and Consumer Loans

Saving and borrowing require comparisons across time. Before using a percentage, we need to know which balance it applies to, how often it is applied, and for how long. The same questions help explain why a low monthly repayment may still produce a high total cost.

These notes provide mathematical tools for examining such comparisons; they do not recommend particular financial products.

Learning Outcomes

After completing this chapter, you should be able to:

- calculate simple interest and maturity value;
- calculate compound amount, compound interest, present value, and effective annual rate;
- explain how inflation affects purchasing power;
- calculate a credit-card finance charge using a simplified average-daily-balance model;
- distinguish interest from reduction of principal in a repayment;
- calculate and interpret a fixed loan repayment; and
- distinguish monthly affordability from total borrowing cost.

7.1 Financial Language

Basic Financial Terms

The **principal**, denoted by P, is the amount initially borrowed or invested.
Interest, denoted by I, is the amount paid for the use of borrowed money or earned by allowing another party to use invested money.
The **interest rate**, denoted by r, is the interest charged or earned during a stated period, expressed as a proportion of the balance to which the rate applies.
The **term** is the length of time for which money is borrowed or invested.
The **maturity value** or **future value**, denoted by A, is the amount held or owed at the end of the stated term.

7.1.1 Rates and Units Must Agree

A percentage must be converted to decimal form before it is used in a formula: $6\% = 0.06$ and $1.5\% = 0.015$.

The units of the rate and the term must also agree. If r is an annual rate, then t must be measured in years. Thus, eight months becomes $t = 8/12 = 2/3$ year.

A rate of 1.5% **per month**, by contrast, may be paired directly with a term measured in months. The number 1.5% means nothing operational until the relevant period has been identified.

The Rate Alone Is Incomplete Information

Before using any advertised rate, ask:

- Is it annual, monthly, or stated for some other period?
- Is it simple or compounded?
- How often is it compounded?
- Are there fees or charges outside the stated rate?
- Is the rate fixed, or may it change?

7.2 Simple Interest

Simple Interest

Simple interest is interest calculated only on the original principal. The interest earned or charged during earlier periods does not itself earn or incur further interest.

If principal P is subject to simple interest at rate r for time t, then

$$I = Prt.$$

The maturity value is the principal plus the interest:

$$A = P + I.$$

Substituting $I = Prt$ gives the equivalent one-step formula

$$A = P(1 + rt).$$

A short simple-interest loan. Suppose ₱12,000 is borrowed for eight months at 6% simple interest per year. Since the rate is annual, $t = 8/12 = 2/3$.

The interest is

$$I = Prt = 12{,}000(0.06)\left(\frac{2}{3}\right) = ₱480.$$

The maturity value is therefore

$$A = P + I = 12{,}000 + 480 = ₱12{,}480.$$

The borrower receives ₱12,000 but repays ₱12,480 at maturity.

The same formula can be rearranged when the unknown quantity is the principal, rate, or term, provided that the corresponding denominator is nonzero:

$$P = \frac{I}{rt}, \qquad r = \frac{I}{Pt}, \qquad t = \frac{I}{Pr}.$$

Recovering an unknown rate. A six-month loan of ₱20,000 incurs ₱800 in simple interest. Since $t = 6/12 = 1/2$, the annual simple-interest rate is $r = I/(Pt) = 800/[20{,}000(1/2)] = 0.08 = 8\%$.

Optional and Non-examinable: Exact and Ordinary Interest. For a term stated in days, some conventions take $t = (\text{number of days})/365$, while others use $t = (\text{number of days})/360$.

These are sometimes called the **exact** and **ordinary** methods, respectively. They produce slightly different answers, so a calculation must state its convention. You will not be examined on this distinction.

7.3 Compound Interest

Simple interest always uses the original principal. Compound interest uses a balance that changes over time.

Compound Interest

Compound interest is calculated on the accumulated balance. After interest is added, that interest becomes part of the balance used in the next calculation.

Suppose the interest rate per compounding period is $i \geq 0$. After one period, the balance is $P(1 + i)$. After two periods, it is $P(1 + i)^2$.

After N periods, the compound amount is

$$A = P(1 + i)^N.$$

If the nominal annual rate is r, interest is compounded m times per year, and the term is t years, then

$$i = \frac{r}{m}, \qquad N = mt,$$

so

$$A = P\left(1 + \frac{r}{m}\right)^{mt}.$$

The compound interest earned or charged is

$$I = A - P.$$

Quarterly compounding. Suppose ₱20,000 is invested for three years at a nominal annual rate of 4.8%, compounded quarterly. Here, $P = 20{,}000$, $r = 0.048$, $m = 4$, and $t = 3$.

Therefore,

$$A = 20{,}000\left(1 + \frac{0.048}{4}\right)^{4(3)} \approx ₱23{,}077.89.$$

The compound interest is $I = A - P = 23{,}077.89 - 20{,}000 = ₱3{,}077.89$.

7.3.1 Why Compound Growth Is Exponential

Under simple interest, the same amount Pr is added during each annual period. Under compound interest, the balance is repeatedly multiplied by the same factor. Repeated addition produces linear growth; repeated multiplication produces exponential growth.

Optional and Non-examinable: Deriving the Compound-Interest Formula. Let A_k denote the balance after k compounding periods. Then $A_1 = P(1+i)$, $A_2 = A_1(1+i) = P(1+i)^2$, and, continuing in the same way, $A_N = P(1+i)^N$.

The formula records the repeated multiplication by $1 + i$.

7.3.2 Present Value

Sometimes the future amount is known and the unknown is the amount required now.

Present Value

The **present value**, denoted by PV, is the amount required now to produce a specified future amount under stated assumptions.

For a present-value calculation, write the compound-growth equation as

$$A = PV(1+i)^N.$$

Dividing by $(1+i)^N$ gives

$$PV = \frac{A}{(1+i)^N}.$$

Saving towards a future amount. How much must be invested now to obtain ₱50,000 after four years at a nominal annual rate of 5%, compounded monthly?

Here, $A = 50{,}000$, $i = 0.05/12$, and $N = 12(4) = 48$.

Thus,

$$PV = \frac{50{,}000}{(1 + 0.05/12)^{48}} \approx ₱40{,}953.55.$$

Under the stated model, investing approximately ₱40,953.55 now produces ₱50,000 after four years.

7.3.3 Nominal and Effective Rates

When interest is compounded more than once a year, the nominal annual rate is not the proportion by which the balance actually grows over the year.

Effective Annual Rate

The **effective annual rate** is the actual proportional increase over one year after accounting for compounding:

$$r_{\text{eff}} = \left(1 + \frac{r}{m}\right)^m - 1.$$

Nominal and effective rates. A nominal annual rate of 6% compounded monthly has an effective annual rate given by

$$r_{\text{eff}} = \left(1 + \frac{0.06}{12}\right)^{12} - 1 \approx 0.061678.$$

The effective annual rate is therefore approximately 6.17%, not exactly 6%.

7.3.4 Inflation and Purchasing Power

Inflation describes a general increase in prices over time. If prices rise, the same nominal amount of money purchases fewer goods and services.

Under a simplified constant-inflation model, a price P after t years at annual inflation rate q becomes

$$A = P(1 + q)^t.$$

A simplified inflation projection. Suppose a basket of goods costs ₱1,000 and prices increase by 4% per year for five years. Under a constant-rate model, $A = 1{,}000(1.04)^5 \approx$ ₱1,216.65.

An amount of ₱1,000 five years later would therefore not purchase the same basket under this model. The calculation is a projection, not a guarantee: actual inflation varies over time and across categories of goods.

Optional and Non-examinable: The Rule of 72. For a moderate annual growth rate of $p\%$, the approximate doubling time is $72/p$ years. At 6%, for example, take $p = 6$, giving $72/6 = 12$ years.

The same estimate provides a quick way to interpret inflation. At a constant inflation rate of 8% per year, prices would take approximately $72/8 = 9$ years to double. This does not mean that every individual price will double: inflation describes a general change in prices, and the rate itself may vary from year to year.

The Rule of 72 is a mental estimate, not an exact formula. It works best for moderate rates under approximately constant compound growth.

7.4 Credit Cards and Consumer Loans

Borrowing provides money now in exchange for future repayments, which include interest and may include fees.

7.4.1 Credit-Card Balances

A credit card does not normally apply interest only to the balance appearing on the final day of a billing period. One common approach uses the average balance held across the days in that period.

Average Daily Balance

If a balance has values $B_1, B_2, \ldots, B_d$ over a billing period of d days, its **average daily balance** is

$$\overline{B} = \frac{B_1 + B_2 + \cdots + B_d}{d}.$$

When the balance remains constant across blocks of days, the same calculation may be written as a weighted mean.

A simplified credit-card finance charge. Suppose a 30-day billing period has the following balances:

Days	Balance
1–10	₱3,000
11–20	₱4,000
21–30	₱2,500

The average daily balance is $\overline{B} = [10(3{,}000) + 10(4{,}000) + 10(2{,}500)]/30 =$ ₱3,166.67.

If the monthly rate is 2%, the finance charge is approximately $0.02(3{,}166.67) =$ ₱63.33.

This is a simplified calculation. An actual card agreement may specify grace periods, fees, exclusions, and a more detailed procedure.

7.4.2 Why Minimum Payments Can Be Deceptive

If interest is added before a payment reduces the balance, only part of the payment reduces the principal.

Separating interest from principal. Consider a balance of ₱10,000 with a monthly interest rate of 2% and a fixed monthly payment of ₱1,000. Assume no new purchases or fees.

Month	Opening balance	Interest	Payment	Principal repaid	Closing balance
1	₱10,000.00	₱200.00	₱1,000.00	₱800.00	₱9,200.00
2	₱9,200.00	₱184.00	₱1,000.00	₱816.00	₱8,384.00
3	₱8,384.00	₱167.68	₱1,000.00	₱832.32	₱7,551.68

Only part of each payment reduces the principal. In the first month, ₱200 pays interest and ₱800 reduces the debt. Reducing the payment while keeping it above the interest charge would slow the reduction of principal.

7.4.3 Fixed-Payment Loans

For a loan of amount P, periodic rate $i > 0$, and n equal end-of-period payments, the payment is

$$R = P\frac{i}{1-(1+i)^{-n}}.$$

If $i = 0$, the payment is simply P/n.

An instalment loan. Suppose ₱60,000 is borrowed at a nominal annual rate of 12%, with equal monthly repayments over one year. Assume monthly compounding and no separate fees.

Here, $P = 60{,}000$, $i = 0.12/12 = 0.01$, and $n = 12$.

The monthly payment is

$$R = 60{,}000\frac{0.01}{1-(1.01)^{-12}} \approx ₱5{,}330.93.$$

The total of the twelve payments is approximately $12(5{,}330.93) = ₱63{,}971.16$, of which approximately ₱3,971.16 is interest. Rounding each monthly payment to the nearest centavo accounts for the slight difference from calculations performed without intermediate rounding.

Optional and Non-examinable: Where the Repayment Formula Comes From. Immediately after the first payment, the balance is $P(1+i) - R$. After the second payment, it is $\big(P(1+i)-R\big)(1+i) - R$.

Continuing produces a geometric sum. Requiring the balance after the final payment to be zero and solving for R gives

$$R = P\frac{i}{1-(1+i)^{-n}}.$$

Thus, the repayment formula follows from requiring the balance to reach zero after n payments.

7.4.4 Advertised Rates and the Cost of Credit

The interest rate is not always the only cost. Application fees, service charges, insurance requirements, penalties, and the timing of payments may change the effective cost of borrowing.

Affordable per Month Does Not Mean Inexpensive

For a fixed principal and interest rate, extending the loan term lowers the periodic payment but usually increases the total interest paid. Compare both:

- the periodic payment; and
- the total amount repaid.

A payment schedule that fits a monthly budget may still be the more expensive option overall.

Optional and Non-examinable: Stated Rate and APR. An **annual percentage rate** attempts to express the annualised cost of credit. Its formal calculation depends on the timing of cash flows and the charges included. Two loans with the same stated interest rate need not have the same effective cost when their fees or payment schedules differ.

In this course, compare the amount received by the borrower with the complete stream of required payments. We shall not attempt the regulatory calculation of APR.

7.5 Exercises

Calculate each amount, then explain what it says about timing or total cost.

1. Calculate the simple interest and maturity value of a ₱7,500 loan for nine months at 8% simple interest per year.
2. A six-month loan of ₱30,000 incurs ₱1,200 in simple interest. Find the annual simple-interest rate.
3. Find the compound amount and compound interest when ₱10,000 is invested for two years at a nominal annual rate of 6%, compounded monthly.
4. Find the effective annual rate corresponding to a nominal annual rate of 12% compounded monthly.
5. Find the present value of ₱25,000 due in three years at an annual rate of 4%, compounded annually.
6. A basket of goods costs ₱2,000. Under a constant annual inflation rate of 5%, estimate its price after three years.
7. During a 30-day billing period, a credit-card balance is ₱2,000 for ten days, ₱3,200 for the next ten days, and ₱2,500 for the final ten days. Calculate the average daily balance and the finance charge at a monthly rate of 1.5%.
8. A loan requires 24 monthly payments of ₱3,100. The borrower receives ₱65,000 and pays no separate fee. Find the total amount repaid and the amount paid above the original principal.
9. A loan of ₱100,000 has a nominal annual rate of 9.6%, compounded monthly, and is repaid through 24 equal monthly payments. Calculate the monthly payment, the total amount repaid, and the amount paid above the original principal. Interpret the final amount as a borrowing cost.

Closing Reflection

Besides the advertised monthly payment, which quantities would you compare before accepting a loan?

Chapter 8

Investments and Home Ownership

Investments exchange money available now for an uncertain future return. A home purchase usually combines an initial cash payment with years of loan repayments. This chapter uses the interest calculations from Chapter 7 to examine both situations, beginning with stocks, bonds, and mutual funds before turning to housing.

Learning Outcomes

After completing this chapter, you should be able to:

- distinguish stocks, bonds, and mutual funds;
- calculate an elementary stock return, dividend yield, bond coupon, and mutual-fund net asset value;
- explain the relationship between diversification and risk;
- calculate a down payment and mortgage amount;
- calculate and interpret a fixed mortgage repayment;
- read a short amortisation schedule;
- distinguish monthly affordability from total ownership cost; and
- state the assumptions and limitations of a financial comparison.

8.1 Stocks, Bonds, and Mutual Funds

Saving and investing are not identical. Saving usually prioritises access and preservation of money. Investing accepts some uncertainty in pursuit of a return.

8.1.1 Stocks

Stock

A **share of stock** represents part ownership of a company. A shareholder may receive a distribution called a **dividend** and may gain or lose money when the market price changes.

Ignoring tax, the net result from an investment in shares may be represented as

$$\text{net result} = \text{sale proceeds} + \text{dividends} - \text{purchase cost} - \text{fees}.$$

Return from shares. An investor buys 50 shares at ₱100 each. The shares are later sold for ₱112 each, and total dividends of ₱2 per share have been received. Suppose all buying and selling fees together amount to ₱150.

The purchase cost is $50(100) =$ ₱5,000. The capital gain is $50(112 - 100) =$ ₱600, and the dividends total $50(2) =$ ₱100. After fees, the net gain is $600 + 100 - 150 =$ ₱550, or $(550/5{,}000) \times 100\% = 11\%$ of the purchase cost.

The calculation describes what happened in this example; it does not predict future returns.

The **dividend yield** compares the annual dividend per share with the market price per share:

$$\text{dividend yield} = \frac{\text{annual dividend per share}}{\text{market price per share}} \times 100\%.$$

Dividend yield is only one component of a return. It does not include a change in the share price or transaction costs.

8.1.2 Bonds

Bond

A **bond** is a debt instrument. The investor lends money to an issuer. The bond states a **face value**, a **coupon rate**, and a **maturity date**. Subject to the terms of the bond and the issuer's ability to pay, the investor receives interest and the face value is repaid at maturity.

For a bond whose annual coupon is calculated from face value,

$$\text{annual coupon payment} = \text{face value} \times \text{coupon rate}.$$

A bond coupon. A bond has a face value of ₱10,000, an annual coupon rate of 5%, and four years remaining to maturity. Its annual coupon payment is $10{,}000(0.05) =$ ₱500.

Over four years, the coupon payments total ₱2,000. If the bond is held to maturity and the issuer meets its obligations, the ₱10,000 face value is also repaid.

This does not mean the investment is risk-free. The issuer may fail to pay, the market value may change before maturity, and inflation may reduce the purchasing power of the payments.

8.1.3 Mutual Funds

Mutual Fund

A **mutual fund** pools money from many investors and holds a portfolio of assets. Each investor owns shares in the fund rather than directly owning a specified fraction of every underlying asset.

One basic measure is the net asset value per share:

$$\text{NAV} = \frac{A - L}{N},$$

where A is the value of the fund's assets, L is its liabilities, and N is the number of fund shares outstanding.

Net asset value. Suppose a fund has assets worth ₱25 million, liabilities of ₱1 million, and 1.2 million outstanding shares. Then NAV = (25,000,000 − 1,000,000)/1,200,000 = ₱20 per share.

8.1.4 Risk and Diversification

Stocks, bonds, and mutual funds expose investors to different combinations of risk, potential return, liquidity, and fees. A fund may provide **diversification** by spreading money across many assets. Diversification can reduce the effect of a poor outcome in one holding, but it cannot eliminate every source of loss.

A Projection Is Not a Promise

Historical growth, an advertised yield, or a modelled return does not guarantee a future result. Before comparing investments, ask:

- How is the return calculated?
- Have fees been deducted?
- How much could be lost?
- How long may the money be unavailable?
- Does the evidence project the future or merely record the past?

8.2 Home Ownership

A home purchase combines an initial payment, long-term borrowing, and continuing expenses. The selling price is therefore not the complete cost of ownership.

8.2.1 Initial Amount Borrowed

If a home has a selling price S and a down payment D, the amount financed is

$$P = S - D.$$

If the down payment is a proportion d of the selling price, then

$$D = dS,$$

and

$$P = S(1 - d).$$

Down payment and initial costs. Suppose a home has a hypothetical selling price of ₱2,000,000 and requires a 20% down payment.

The down payment is $D = 0.20(2{,}000{,}000) =$ ₱400,000. The amount financed is $P = 2{,}000{,}000 - 400{,}000 =$ ₱1,600,000. If other initial fees in this simplified example total 3% of the selling price, they amount to $0.03(2{,}000{,}000) =$ ₱60,000.

The initial cash required is therefore ₱460,000. The 3% figure is an assumption for this example, not a statement of a universal or current fee.

8.2.2 Mortgage Repayments

A fixed-payment mortgage uses the repayment formula introduced in Chapter 7:

$$R = P\frac{i}{1-(1+i)^{-n}}.$$

A fixed mortgage repayment. Suppose ₱1,600,000 is financed for 15 years at a nominal annual rate of 6%, with equal monthly payments. Assume that the rate remains fixed and ignore fees and insurance.

Here, $i = 0.06/12 = 0.005$ and $n = 15(12) = 180$.

The monthly payment is

$$R = 1{,}600{,}000\frac{0.005}{1-(1.005)^{-180}} \approx ₱13{,}501.71.$$

The total of 180 payments is approximately $180(13{,}501.71) = ₱2{,}430{,}307.80$. Of this amount, approximately $2{,}430{,}307.80 - 1{,}600{,}000 = ₱830{,}307.80$ is interest. A borrower should compare the monthly payment with both the total repayment and the total interest.

8.2.3 Amortisation

Amortisation

Amortisation is the gradual repayment of a debt through scheduled payments. Each payment is divided between interest and reduction of the principal.
An **amortisation schedule** records the opening balance, interest, payment, principal reduction, and closing balance for each period.

For the mortgage above, the first two rows of its amortisation schedule, rounded to centavos, are:

Month	Opening balance	Interest	Payment	Principal repaid	Closing balance
1	₱1,600,000.00	₱8,000.00	₱13,501.71	₱5,501.71	₱1,594,498.29
2	₱1,594,498.29	₱7,972.49	₱13,501.71	₱5,529.22	₱1,588,969.07

As the balance falls, the interest portion falls and more of each fixed payment reduces the principal.

8.2.4 The Repayment Is Not the Entire Cost

Possible continuing costs include:

- property-related taxes and charges;
- insurance;
- association or condominium fees;
- utilities;

- repairs and maintenance; and
- transport costs associated with the location.

Which items apply and how much they cost depend on the property, the agreement, and the applicable law. Check them instead of assuming them.

Renting Versus Buying Is a Modelling Question

There is no universal mathematical rule that buying is always better than renting, or the reverse. A useful comparison depends on assumptions about:

- how long the person expects to remain in the home;
- interest rates and fees;
- rent and property-price changes;
- maintenance and insurance;
- alternative uses of the down payment; and
- uncertainty in income and personal circumstances.

8.3 Making a Sound Financial Comparison

For the examples in these notes, we use the following order:

1. Identify the money received, paid, borrowed, or invested at the beginning.
2. Identify each rate and the period to which it applies.
3. Determine whether the calculation uses simple or compound interest.
4. Record the number and timing of payments.
5. Include fees and other unavoidable costs.
6. Calculate both the periodic amount and the total amount.
7. State assumptions about inflation, future returns, and changing rates.
8. Consider risks that are not captured by the calculation.

Cash price versus instalments. Suppose an item costs ₱24,000 in cash or may be purchased using twelve monthly payments of ₱2,150 with an additional one-time fee of ₱500.

The instalment option costs $12(2{,}150) + 500 =$ ₱26,300. The difference from the cash price is $26{,}300 - 24{,}000 =$ ₱2,300.

Under the stated assumptions, instalments cost ₱2,300 more. The buyer must decide whether spreading the payments is worth that additional cost.

Financial Education, Not Individual Financial Advice

These calculations are hypothetical. They omit many fees, tax rules, contractual terms, risks, and personal circumstances. Read the current terms before entering a financial agreement, and seek qualified advice when needed.

8.4 Exercises

In each comparison, state the assumptions, costs, and risks you have considered.

1. An investor buys 40 shares at ₱80 each and later sells them for ₱92 each. During the holding period, dividends total ₱1.50 per share, while all fees total ₱100. Calculate the net gain.
2. A share pays an annual dividend of ₱3.60 and has a market price of ₱120. Calculate its dividend yield.
3. A bond has a face value of ₱20,000 and an annual coupon rate of 4.5%. Find the annual coupon payment and the total coupon payments over five years.
4. A mutual fund has assets of ₱10 million, liabilities of ₱1 million, and 450,000 outstanding shares. Calculate its NAV per share.
5. Explain why diversification may reduce risk without eliminating it.
6. A home has a selling price of ₱3,000,000 and requires a 15% down payment. Find the down payment and mortgage amount.
7. For the amortisation table above, explain why the interest component falls while the principal component rises.
8. An item costs ₱18,000 in cash or twelve monthly payments of ₱1,600 plus a fee of ₱300. Find the total instalment cost and its difference from the cash price.
9. Use the ₱2,550,000 mortgage amount from Question 6. If the nominal annual rate is 6%, compounded monthly, and the term is 20 years, calculate the fixed monthly repayment and the total of the scheduled repayments. State one important cost not included in this calculation.
10. The instalment option in Question 8 costs more in total than the cash option. Give two assumptions or circumstances that should be considered before concluding that paying cash is the better decision for a particular buyer.

Closing Reflection

Which uncertain assumption would have the greatest effect on a long-term investment or housing comparison?

Chapter 9

The Mathematics of Graphs

A list of organisms may record their species, age, mass, or location. Such a list describes the organisms one at a time. It does not immediately answer a different kind of question: which organisms interacted with which? In that question, the relationships are as important as the organisms themselves.

Graph theory provides a mathematical language for relationships. It allows us to represent individuals and infectious contact, species and feeding relationships, habitat patches and wildlife corridors, or field sites and the routes between them. The biological meanings differ, but the mathematical structure may be the same.

Learning Outcomes

After completing this chapter, you should be able to:

- identify vertices, edges, adjacent vertices, neighbours, and degrees;
- identify paths and connected components;
- interpret directed, weighted, tree, and bipartite graph models;
- construct and interpret a proper vertex colouring in an applied setting; and
- explain the modelling choices and limitations behind a biological graph.

9.1 From Biological Relationships to Graphs

Suppose six animals are being observed. Their measured characteristics fit naturally into a table. If we instead ask which pairs had sufficiently close contact for a pathogen to pass between them, the central information consists of pairs. A network picture can make those relationships easier to see.

Graph

A **graph** consists of a set of **vertices** and a set of **edges**. Vertices represent the objects under study, and edges represent the selected relationship between pairs of objects.

Unless stated otherwise, the graphs in this chapter have no loops and no repeated edges between the same pair of vertices.

Before drawing a graph, we must therefore ask: what should the vertices and edges represent? The answer depends on the question. A line between two animals might mean physical contact, shared feeding ground, genetic similarity, or something else. The drawing alone does not decide its meaning.

9.1.1 The Same Structure in Different Settings

Consider the graph below.

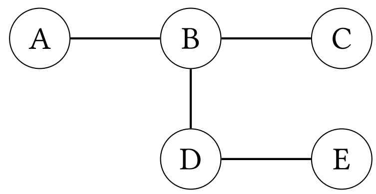


Figure 9.1: One abstract graph. Its interpretation depends on what the vertices and edges have been chosen to represent.

If the vertices are people and the edges are infectious contacts, the graph is a contact network. If the vertices are forest patches and the edges are wildlife corridors, it is a habitat network. If the vertices are species and the edges are ecological interactions, it is an ecological network. Mathematically, all three models have the same arrangement of vertices and edges.

This is an example of **abstraction**. We set aside details that do not matter to the question and retain the structure that does. The abstraction is useful precisely because a conclusion about the graph may apply to several different settings. It is also limited: information omitted from the graph cannot later be recovered from the drawing.

Two populations may contain the same number of individuals and the same number of recorded interactions yet have differently arranged edges. One may form a single connected network while the other divides into separate groups. Counts alone do not describe that difference.

9.2 A Contact Network

In the following hypothetical network, each vertex represents an individual. An edge records sufficiently close contact during a stated period for direct transmission to have been possible. It does not assert that transmission actually occurred.

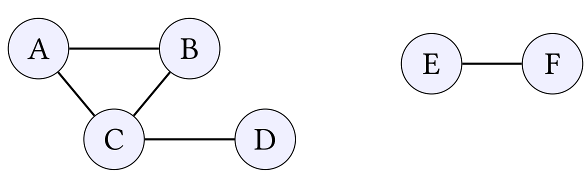


Figure 9.2: A hypothetical undirected contact network with two connected components.

The graph is **undirected**: contact between A and B is treated as the same relationship as contact between B and A. Vertices A and B are **adjacent** because an edge joins them. Each is a **neighbour** of the other. A is not adjacent to D.

Degree

The **degree** of a vertex in an undirected graph is the number of edges incident with it. Equivalently, in a graph without repeated edges or loops, it is the number of neighbours of that vertex.

In Figure 9.2, C has degree 3, A and B have degree 2, and D, E, and F have degree 1. C therefore has the most recorded contacts. This alone does not make C the most important individual for disease spread: the graph omits contact duration and timing, susceptibility, infectiousness, and unrecorded contacts.

9.2.1 Paths and Connectedness

A **path** follows consecutive edges from one vertex to another without repeating a vertex. In Figure 9.2, A–C–D is a path from A to D. In the contact model, it represents a possible chain of transmission, not evidence that transmission actually followed that chain.

Two vertices are **connected** if a path joins them. A graph is connected if every pair of vertices is connected. A **connected component** is a maximal connected part of a graph.

Figure 9.2 has two components: one contains A, B, C, and D; the other contains E and F. Thus, A and D are connected in the static graph, whereas A is disconnected from E and F. A path identifies a candidate transmission chain, but the timing of contacts must also permit transmission. A missed interaction could also join the components.

9.3 Direction, Weights, and Biological Meaning

Some relationships cannot be represented adequately by an undirected edge. If a graph records who infected whom, or which organism is consumed by which, direction matters.

9.3.1 A Directed Food Web

In the hypothetical food web below, an arrow points from food to consumer. The edge grasshopper → frog therefore means that the frog consumes the grasshopper.

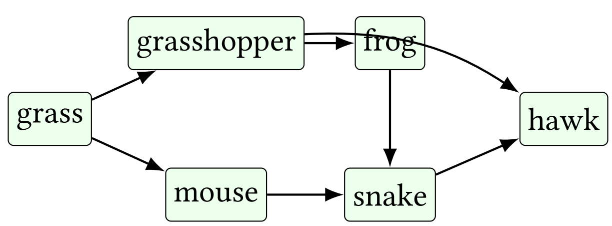


Figure 9.3: A hypothetical directed food web in which arrows point from food to consumer.

Directed Graph

A **directed graph**, or **digraph**, has edges with specified directions. The **indegree** of a vertex counts arrows entering it, and the **outdegree** counts arrows leaving it.

The hawk has indegree 2 in Figure 9.3, because it receives arrows from the grasshopper and snake. Grass has outdegree 2. These numbers depend on the convention for the arrows. If arrows instead pointed from consumer to food, the represented feeding relationships would be unchanged, but each vertex's indegree and outdegree would be interchanged. Every directed food web should therefore state what an arrow means.

Removing a species from the drawing can reveal which represented pathways are lost, but it cannot by itself predict the resulting ecosystem. Population sizes, interaction strengths, behaviour, and environmental change have been omitted.

9.3.2 Weighted Graphs

An edge may carry more information than mere presence or absence. A **weighted graph** assigns a numerical weight to each edge. Depending on the model, a weight might represent distance, travel time, interaction frequency, energetic cost, or estimated transmission probability.

Suppose the vertices in Figure 9.4 represent a field station S and three habitat patches. Edge weights give walking distance in kilometres.

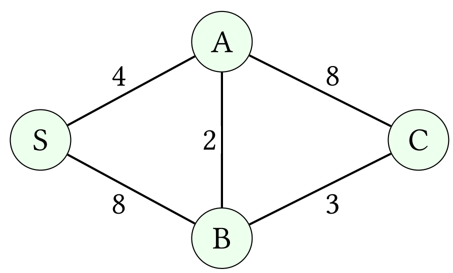


Figure 9.4: A weighted graph whose edge weights are distances in kilometres.

The route S–A–C has total length $4 + 8 = 12$ km, while S–B–C has length $8 + 3 = 11$ km. The route S–A–B–C uses more edges but has total length $4 + 2 + 3 = 9$ km. The path with the fewest edges need not have the least total weight.

9.3.3 A Bipartite Pollination Network

A graph is **bipartite** if its vertices can be divided into two groups so that every edge joins one group to the other. A plant–pollinator network is naturally bipartite when one group contains plants, the other contains pollinators, and an edge records an observed pollination interaction. No edge then joins two plants or two pollinators.

Such a graph can show that a pollinator visited many plant species or that a plant was observed with only one pollinator. It does not, without further data, measure the effectiveness of each visit or prove that an organism with low degree will disappear.

9.4 Trees and Phylogenetic Diagrams

A tree is a useful graph when the relationships branch without later joining again.

Tree

A **tree** is a connected graph in which exactly one path joins any two distinct vertices.

Phylogenetic relationships are often represented by trees. Tips or leaves may represent observed taxa, while internal branching points represent common ancestral lineages in the model.

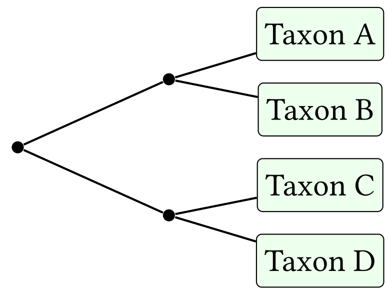


Figure 9.5: A simplified phylogenetic tree. A and B share one recent branching point, while C and D share another.

The branching pattern, rather than the left-to-right order of the labels, is the important information. Rotating the two branches around an internal point does not change the represented relationships. Branch lengths have no numerical meaning unless the diagram says that they do. Constructing a phylogeny from biological evidence is a much deeper problem than reading the relationships shown by a completed tree.

9.5 Graph Colouring in a Biological Laboratory

Suppose several laboratory procedures must be assigned to sessions. Join two procedures by an edge when they cannot occur in the same session because they require the same limited equipment or personnel. A colour can then represent a session.

Proper Vertex Colouring

A **proper vertex colouring** assigns colours to vertices so that adjacent vertices receive different colours.

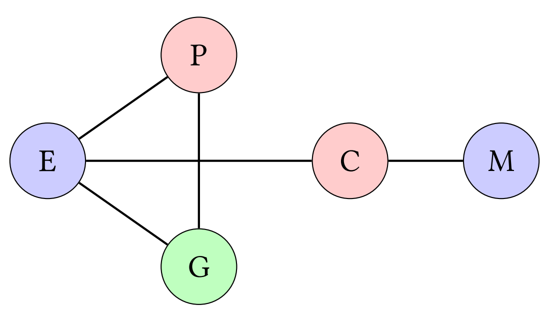


Figure 9.6: A proper three-colouring of a hypothetical laboratory conflict graph. Vertices E, P, G, C, and M represent procedures.

The triangle E–P–G–E forces E, P, and G to receive three different colours, so two colours cannot suffice. Figure 9.6 gives a proper colouring with three colours. Here a colour represents a laboratory session, so the colouring provides a schedule in which conflicting procedures are separated. The same idea can assign treatments to neighbouring plots when adjacent plots must receive different treatments. The meaning of a colour comes from the application, not from graph theory itself.

9.6 What the Graph Leaves Out

A graph begins with modelling choices. In a contact network, what duration or distance counts as contact? Should a repeated interaction create a heavier edge? Should an edge point from a source of infection to a recipient? Does the observation period combine interactions that occurred at very different times?

Changing these decisions can change the graph and the conclusions drawn from it. A graph is not a neutral photograph of reality. It is a deliberately simplified representation built for a question. Its value depends on whether the chosen vertices, edges, directions, and weights preserve the information needed to answer that question.

Structure Is Not the Whole Biological Story

A high degree, a connected component, or a short weighted path can identify an important feature of a graph. None of these graph features automatically establishes biological importance, causation, or an intervention priority. Such conclusions require biological knowledge, suitable data, and attention to what the graph omits.

9.7 Exercises

For every biological graph, state what its vertices and edges represent before interpreting it.

1. A contact network has vertex set $\{A, B, C, D, E, G\}$ and edge set $\{AB, AC, BC, CD, DE\}$. List the neighbours and degree of each vertex, identify its connected components, and give one path from A to E.
2. The graph in Figure 9.1 is used first for infectious contacts and then for wildlife corridors. State one conclusion that is mathematically valid in both interpretations and one question that cannot be answered without knowing the interpretation.
3. A directed food web contains the arrows $P \to I$, $P \to R$, $I \to F$, $I \to B$, $R \to H$, and $F \to H$, where arrows point from food to consumer. Find the indegree and outdegree of every vertex. Interpret the arrow $I \to F$ and the indegree of H.
4. A habitat network has patches A, B, C, D, E and corridors AB, BC, DE. Identify its connected components. Give one additional corridor that would make the whole network connected, and explain what the new edge represents biologically.
5. A weighted habitat network has routes S–A–T with weights 5 and 7, S–B–T with weights 8 and 3, and S–A–B–T with weights 5, 1, and 3. Find the total weight of each route. Which has the least total weight, and why would choosing the route with the fewest edges fail?

6. In Figure 9.5, which pair of taxa shares the upper branching point? If the positions of A and B were exchanged by rotating the branches around that point, would the represented relationship change? Explain.
7. A conflict graph for five laboratory activities has edges AB, AC, BC, BD, CE. Find a proper colouring using as few colours as possible and justify why fewer colours cannot work.
8. Researchers want to model interactions between three flowering plants and four pollinator species. State what the vertices and edges should represent, whether the graph should be directed or undirected, whether weights might be useful, and one important kind of information the graph would omit.
9. Two disease-contact graphs each have 20 vertices and 30 edges. Must an outbreak behave the same way in both? Explain using at least two structural features that the totals do not determine.
10. A species has the highest total degree (indegree plus outdegree) in a directed food web. Does this fact alone show that it is the most important species in the ecosystem? Give a graph-based interpretation of the high total degree and two biological considerations that the degree does not capture.

Closing Reflection

Think of a biological question whose main information concerns relationships. What should the vertices and edges represent, and what would your graph deliberately leave out?

Appendix A

Solutions to Exercises

The solutions in this appendix are model answers. Other well-justified answers are possible.

A.1 Chapter 1: Mathematics in Our World

1. A useful model might include distance, transport, traffic, weather, waiting time, and road closures. The vehicle's colour could be omitted because it does not materially affect journey time.
2. The next three terms are 34, 55, and 89. The corresponding ratios are

$$\frac{34}{21} \approx 1.6190, \qquad \frac{55}{34} \approx 1.6176, \qquad \frac{89}{55} \approx 1.6182.$$

   These support the conjecture that the ratios approach φ, but finitely many examples cannot prove a statement about the entire sequence.
3. One possible measurement is the number or area of visible yellow lesions. It captures one sign of diabetic retinopathy but may overlook vessel damage, image quality, lesion type, or location.
4. For a claim that a shell follows the golden ratio, we would need defined measurements, a reliable procedure, a suitable sample, and a stated tolerance for closeness to φ. Visual resemblance alone is not sufficient evidence.
5. Mathematics extends beyond calculation. A navigation application represents roads as a network and uses several mathematical tools to select a route; the Fibonacci and golden-ratio examples use numbers, diagrams, and logical relationships to investigate natural patterns. These illustrate practical, intellectual, and aesthetic dimensions of mathematics.
6. The developers should disclose the test data, error rates, and loss of performance on blurry images. They should state that the system supports screening rather than diagnosis and that uncertain cases require clinical review. Integrity requires reporting assumptions, uncertainty, and known limitations—not only a favourable overall accuracy.

A.2 Chapter 2: The Language of Mathematics

1. (a) $3x + 5$ is an **expression**.
   (b) $3x + 5 = 20$ is an **open sentence** because its truth depends on the value of x.
   (c) $3(5) + 5 = 20$ is a **closed sentence**, and it is true.
   (d) $\forall x \in \mathbb{R},\ x^2 \geq 0$ is a **closed sentence**, and it is true.

2. (a) $P > 500$.
   (b) $0 \leq p \leq 1$.
   (c) $s \in E$.
   (d) The sets A and B have no elements in common.
   (e) For every x in D, applying f twice returns x.
   (f)
   $$r = \frac{N_2 - N_1}{t_2 - t_1}.$$

3. Division and multiplication have equal priority and are performed from left to right:

   $$20 - 12 \div 3 \times 2 = 20 - 4 \times 2 = 20 - 8 = 12.$$

4. Here
   $$M = \{\text{bat}, \text{whale}\}, \qquad F = \{\text{bat}, \text{eagle}\}.$$

   The bat belongs in the overlap, the whale in the M-only region, the eagle in the F-only region, and the penguin outside both circles. Therefore,

   $$M \cap F = \{\text{bat}\}, \qquad M \cup F = \{\text{bat}, \text{eagle}, \text{whale}\}.$$

5. One suitable relation is

   $$R = \{(\text{cat}, \text{mammal}), (\text{dog}, \text{mammal}), (\text{eagle}, \text{bird})\}.$$

   It is also a function because each element of A is paired with exactly one element of B.

6.

P	Q	$\neg Q$	$P \wedge \neg Q$
T	T	F	F
T	F	T	T
F	T	F	F
F	F	T	F

The implication $P \Rightarrow Q$ says:

> If the organism is a dog, then it is a mammal.

Its converse says:

> If the organism is a mammal, then it is a dog.

The implication is true, but its converse is false.

7. (a) At least one organism in the sample does not exhibit the trait.
   (b) Every patient has a temperature below 38°C.

8. Addition on $\mathbb{Z}$ and multiplication on $\mathbb{R}$ are binary operations.

   Subtraction on the positive integers is not a binary operation because, for example, $2 - 3 = -1$, which is not positive. Division on $\mathbb{R}$ is not a binary operation because division by zero is undefined.

9. $$5 \star (-2) = 1, \qquad (1 \star 3) \star 2 = 11, \qquad 1 \star (3 \star 2) = 15.$$

   Because the final two results are different, $\star$ is not associative.

10. For example, let a set consist of the specimens included in a study. Let a relation record which traits each specimen exhibits, and let a function assign each specimen its recorded mass.

11. We must identify the specimens under discussion, the quantity being measured and its unit, and what "one measurement" means. For example, does it mean one reading taken at a specified time or one recorded variable?

A.3 Chapter 3: Problem-Solving Strategies and Formulas

1. **Understand the problem:** determine what is given and what must be found. **Devise a plan:** select a connection or method that may lead to the answer. **Carry out the plan:** execute and justify the steps. **Look back:** check the result, its assumptions, and its meaning.

2. Pairing the first and last terms produces the same sum as pairing the second and second-to-last terms. This turns a long addition into 50 equal pairs, making the total easy to calculate.

3. Substitution checks whether the result satisfies the original equation, so it belongs to **Look back**.

4. **Understand:** each handshake corresponds to one unordered pair of people. **Devise a plan:** count the new handshakes contributed as the people are considered one at a time. **Carry out:** the counts are
   $$11 + 10 + \cdots + 1 = \frac{11}{2}(11 + 1) = 66.$$

   Look back: the calculation $12(11)/2 = 66$ counts each pair once and confirms the result.

5. Here $t_1 = 5$, $t_n = 50$, and $d = 3$. Using $t_n = t_1 + (n-1)d$,
   $$50 = 5 + 3(n - 1).$$

   Hence $45 = 3(n - 1)$, so $n = 16$.

6. $$S = \frac{16}{2}(5 + 50) = 8(55) = 440.$$

7. Writing an arithmetic series forwards and backwards produces n pairs, each equal to $t_1 + t_n$. Thus $2S = n(t_1 + t_n)$, from which $S = \frac{n}{2}(t_1 + t_n)$. The formula records this pairing argument.
8. Memorisation saves time and lets us keep our attention on the problem at hand. It may also make a relationship easier to implement in a computer. We still need to understand the formula's assumptions.
9. If $a = 0$, the equation is not quadratic. The derivation also divides by a, which would be invalid when $a = 0$.
10. Here $a = 1$, $b = 1$, and $c = 1$, so $b^2 - 4ac = 1 - 4 = -3$. Since the discriminant is negative, the equation has no real solutions.

A.4 Chapter 4: Reasoning, Evidence, and Proof

1. A conclusion may be reasonable when several relevant and representative observations support the same pattern. It remains uncertain because an unexamined case may provide a counterexample or another explanation.
2. The premises are "All mammals are vertebrates" and "A dolphin is a mammal". The conclusion is "A dolphin is a vertebrate". The reasoning is deductive because the conclusion necessarily follows if both premises are true.
3. Repeated observations that a treated group recovers more quickly than a comparison group may support the hypothesis that the treatment is effective. Chance, bias, confounding variables, or flaws in the study may still provide other explanations.
4. Evidence shows that a claim holds in the cases examined and may make the claim plausible. A proof establishes, through definitions and valid reasoning, that it holds in every case covered by the statement.
5. Suppose, for contradiction, that N is the largest integer. Then $N + 1$ is also an integer and $N + 1 > N$. This contradicts the assumption that N is the largest integer. Therefore, there is no largest integer.
6. Multiplication by 10 shifts the decimal point:

   $$10(0.9999\ldots) = 9.9999\ldots.$$

   The part after the decimal point is still $0.9999\ldots$, so $9.9999\ldots = 9 + 0.9999\ldots$.
7. The examples suggest that the sum of an even integer and an odd integer is odd. Three cases cannot establish the claim for every pair of integers.
8. Since m and n are integers, $m + n$ is an integer. The expression $2(m + n) + 1$ therefore has the defining form $2k + 1$ of an odd integer.
9. Let $2m + 1$ and $2n + 1$ be arbitrary odd integers. Their sum is

   $$(2m + 1) + (2n + 1) = 2(m + n + 1).$$

   Since $m + n + 1$ is an integer, the sum has the form $2k$ and is therefore even.

10. For $n = 1$, both sides equal 1, so the base case is true. Assume that

$$1 + 2 + \cdots + k = \frac{k(k+1)}{2}$$

for some positive integer k. Then

$$\begin{aligned} 1 + 2 + \cdots + k + (k+1) &= \frac{k(k+1)}{2} + (k+1) \\ &= \frac{(k+1)(k+2)}{2}. \end{aligned}$$

This is the required formula with $n = k + 1$. The base case and inductive step show that the formula holds for every positive integer n.

11. The number 9 is odd but not prime. The original statement applies to every odd integer, so one odd integer without the claimed property disproves it.

A.5 Chapter 5: Managing and Describing Data

1. The four planted seeds are the observations. Daily light exposure and height are numerical variables; leaf colour and germination status are categorical variables. Seed ID is the identifier. For S04, height was not measured, and leaf colour is not applicable because no seedling developed.
2. The numbers 1 and 2 are merely labels for categories. Arithmetic on those labels has no biological meaning: treatment 2 is not "twice" treatment 1.
3. For example, the population might be all first-year biology students at a university, while a sample might be 50 of those students selected for a survey. The population is the complete group of interest; the sample is the subset actually observed.
4. A larger sample reduces random sampling variation, but it does not repair a systematic bias. If the selection method repeatedly excludes part of the population, collecting more observations by the same method reproduces the same problem.
5. For 4, 5, 5, 6, 10,

$$\bar{x} = 6 \text{ days}, \qquad \text{median} = 5 \text{ days}, \qquad \text{mode} = 5 \text{ days}, \qquad \text{range} = 6 \text{ days}.$$

The squared deviations from the mean sum to 22 days^2, so the sample variance and sample standard deviation are

$$s^2 = \frac{22}{4}\,\text{days}^2 = 5.5\,\text{days}^2, \qquad s = \sqrt{5.5}\,\text{days} \approx 2.35\,\text{days}.$$

The median is less affected by the relatively large value 10.

6. The trays contain $0.90(10) = 9$ and $0.60(40) = 24$ germinated seeds. Therefore,

$$\frac{9 + 24}{10 + 40} = \frac{33}{50} = 0.66,$$

giving a combined germination percentage of 66%.

7. The rule identifies a value that warrants attention; it does not identify the cause. The value may be an error or a genuine, scientifically important observation, so check how it was obtained and recorded before deciding what to do with it.

8. A histogram represents intervals of a numerical scale, and adjacent intervals share boundaries; its bars therefore ordinarily touch. A bar chart represents distinct categories, which are conventionally separated by gaps.

9. The variables are daily light exposure, measured in hours, and seedling height, measured in centimetres. The plot shows a generally positive but imperfect association: seedlings receiving more light tend to be taller.

10. A suitable data dictionary might state that `specimen_id` is a unique identifier stored as text; `species` records the accepted species label from a stated list; and `mass_g` records measured mass in grams as a nonnegative numerical value. The precise allowed species labels and the symbol used for a missing mass should also be recorded.

11. A height of zero would assert that a seedling developed and was measured at exactly zero centimetres. Here no seedling developed. The height should be recorded using the dataset's designated code for an unavailable value, and a separate note or status variable should record that germination did not occur.

12. The project should not collect names, and it should omit medical histories because neither is needed for the stated question. Access should be restricted, and participants should be told what information is collected, why it is needed, and how it will be protected.

A.6 Chapter 6: Probability, Association, and Prediction

1. $$z = \frac{45 - 39}{3} = 2.$$

   The measurement is two standard deviations above the population mean.

2. Standardising gives
   $$z = \frac{130 - 100}{15} = 2.$$
   Hence $P(X > 130) = P(Z > 2) \approx 0.0228$. Under the model, about 2.28% of observations exceed 130.

3. For example, the population distribution may not be approximately normal, or the sample may be biased and therefore fail to represent the intended population. Measurement error or dependence among observations could also undermine the interpretation.

4. Pearson's correlation measures only linear association. A strong curved pattern can therefore have correlation near zero; a U-shaped pattern is one example.

5. **Interpolation** predicts within the range of observed explanatory values; **extrapolation** predicts beyond that range. The immunisation example used extrapolation because 80% immunisation lay below the observed values, which ranged from 88% to 92%.

6. $$\hat{y} = 4 + 2.5(6) = 19.$$

   The predicted seedling height is 19 cm. The slope means that each additional hour of daily light is associated with an increase of 2.5 cm in predicted height under the fitted model. This interpretation applies only over the range for which the line was fitted and does not establish that extra light causes the increase.

7. The association may be affected by confounding variables such as health, stress, study habits, or workload. Direction of influence is also unclear, and a population-level trend does not guarantee that changing one individual's sleep will cause the predicted change in score.

8. One variable is usual journey time to campus. Students who already favour the proposal may be more likely to respond, creating voluntary-response bias. Exact home addresses or travel histories could identify individuals, so the study should collect only the location detail needed and protect it appropriately.

A.7 Chapter 7: Interest, Credit, and Consumer Loans

1. Since $t = 9/12 = 3/4$,
   $$I = 7{,}500(0.08)\left(\frac{3}{4}\right) = ₱450.$$
   Therefore,
   $$A = 7{,}500 + 450 = ₱7{,}950.$$

2. Since $t = 6/12 = 1/2$,
   $$r = \frac{1{,}200}{30{,}000(1/2)} = 0.08.$$
   The annual simple-interest rate is 8%.

3. $$A = 10{,}000\left(1 + \frac{0.06}{12}\right)^{24} \approx ₱11{,}271.60.$$
   The compound interest is approximately ₱1,271.60.

4. $$r_{\text{eff}} = \left(1 + \frac{0.12}{12}\right)^{12} - 1 \approx 0.126825.$$
   The effective annual rate is approximately 12.68%.

5. $$PV = \frac{25{,}000}{(1.04)^3} \approx ₱22{,}224.91.$$

6. $$A = 2{,}000(1.05)^3 = ₱2{,}315.25.$$

7. $$\overline{B} = \frac{10(2{,}000) + 10(3{,}200) + 10(2{,}500)}{30} = ₱2{,}566.67.$$

The finance charge is approximately

$$0.015(2{,}566.67) = ₱38.50.$$

8. The total amount repaid is

$$24(3{,}100) = ₱74{,}400.$$

The amount above the original principal is

$$74{,}400 - 65{,}000 = ₱9{,}400.$$

9. The monthly rate is $i = 0.096/12 = 0.008$, and $n = 24$. Hence

$$R = 100{,}000\frac{0.008}{1-(1.008)^{-24}} \approx ₱4{,}596.05.$$

Using the unrounded repayment, the total is approximately ₱110,305.26, so approximately ₱10,305.26 is paid above the original principal. Under the stated assumptions and with no additional fees, this difference is the cost of borrowing represented by the repayment schedule.

A.8 Chapter 8: Investments and Home Ownership

1. The purchase cost is $40(80) = ₱3{,}200$. The capital gain is $40(92 - 80) = ₱480$, and dividends total $40(1.50) = ₱60$. After fees, the net gain is $480 + 60 - 100 = ₱440$.

2. $$\frac{3.60}{120} \times 100\% = 3\%.$$

3. The annual coupon payment is

$$20{,}000(0.045) = ₱900.$$

Over five years, the coupon payments total

$$5(900) = ₱4{,}500.$$

4. $$\text{NAV} = \frac{10{,}000{,}000 - 1{,}000{,}000}{450{,}000} = ₱20$$

per share.

5. Diversification spreads money across several holdings, reducing the effect of a poor result in any one of them. The holdings may still be exposed to common market, economic, or institutional risks, so loss remains possible.

6. The down payment is

$$0.15(3{,}000{,}000) = ₱450{,}000.$$

The mortgage amount is

$$3{,}000{,}000 - 450{,}000 = ₱2{,}550{,}000.$$

7. Interest is calculated from the outstanding balance. As the balance falls, the interest charged falls. With a fixed payment, more of each later payment therefore goes towards principal.

8. The instalment cost is
$$12(1{,}600) + 300 = ₱19{,}500.$$
It exceeds the cash price by
$$19{,}500 - 18{,}000 = ₱1{,}500.$$

9. Here $P = 2{,}550{,}000$, $i = 0.06/12 = 0.005$, and $n = 20(12) = 240$. Thus
$$R = 2{,}550{,}000\frac{0.005}{1-(1.005)^{-240}} \approx ₱18{,}268.99.$$
Using the unrounded repayment, the scheduled repayments total approximately ₱4,384,558.08. This assumes that the nominal rate remains fixed and excludes taxes, insurance, association dues, maintenance, and transaction charges.

10. Relevant considerations include whether paying cash would exhaust the buyer's emergency savings, whether the buyer could earn a return on money retained, whether the instalment agreement contains further charges or penalties, and whether future income is sufficiently reliable to meet the payments. The smaller stated total is important, but it is not the only factor in the decision.

A.9 Chapter 9: The Mathematics of Graphs

1. The neighbours are

Vertex	Neighbours	Degree
A	B, C	2
B	A, C	2
C	A, B, D	3
D	C, E	2
E	D	1
G	none	0

The components are $\{A, B, C, D, E\}$ and $\{G\}$. One path from A to E is A–C–D–E.

2. In either interpretation, B has degree 3, the graph is connected, and every path between A and E passes through B and D. The biological meaning of removing an edge cannot be determined without the interpretation: it might prevent a possible transmission or remove a movement corridor.

3. The degrees are

Vertex	Indegree	Outdegree
P	0	2
I	1	2
R	1	1
F	1	1
B	1	0
H	2	0

   The arrow $I \to F$ means that F consumes I. The indegree 2 of H means that H consumes two represented food sources, R and F.

4. The components are $\{A, B, C\}$ and $\{D, E\}$. Any corridor joining a vertex in the first component to a vertex in the second will connect the network; CD is one example. Biologically, the new edge represents a usable movement corridor between the two previously separated groups of habitat patches.

5. The route totals are $5 + 7 = 12$, $8 + 3 = 11$, and $5 + 1 + 3 = 9$, respectively. S–A–B–T has the least total weight even though it uses three edges. Counting edges ignores their unequal weights.

6. A and B share the upper branching point. Exchanging their displayed positions around that point does not change the relationship: both remain attached to the same branching point. The branching structure matters, not which label is drawn above the other.

7. The triangle formed by A, B, and C requires three different colours, so at least three are necessary. One proper colouring is $A = 1$, $B = 2$, $C = 3$, $D = 1$, and $E = 1$. Thus, three colours suffice and fewer cannot work.

8. Use one group of vertices for the three plants and another for the four pollinators. Join a plant to a pollinator when the selected interaction is observed. An undirected bipartite graph is suitable if only the pairing matters. Weights could record visit frequency or estimated pollination effectiveness. The model may omit abundance, time of day, seasonal change, unsuccessful visits, and environmental conditions.

9. No. The two graphs may have different connected components, degree distributions, paths, or concentrations of edges around a few individuals. They may therefore provide different possible routes for transmission even though their numbers of vertices and edges agree. Timing, direction, and intensity of contact could create further differences not represented by an unweighted graph.

10. High total degree means that the species has more incoming and outgoing feeding links in total than any other vertex under the stated arrow convention. It does not measure interaction strength, population abundance, functional redundancy, susceptibility to disturbance, or the consequences of removing the species. Those considerations require further evidence.

Bibliography

[1] Commission on Higher Education. Mathematics in the Modern World: Preliminaries. Course curriculum issued under CHED Memorandum Order No. 20, series of 2013, 2013.

[2] Edsger W. Dijkstra. A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: His Life, Work, and Legacy*, pages 287–290. Association for Computing Machinery, 2022. doi: 10.1145/3544585.3544600.

[3] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2): 100–107, 1968. doi: 10.1109/TSSC.1968.300136.

[4] DevTex. Why navigation apps always find the perfect route? YouTube video, n.d. URL https://www.youtube.com/watch?v=YwVSpPZ5goc.

[5] Cristóbal Vila. Nature by numbers. YouTube video, 2010. URL https://www.youtube.com/watch?v=kkGeOWYOFoA.

[6] Euclid. Elements, book VI, definition 3. Web edition by David E. Joyce, Clark University, 1997. URL https://math.clarku.edu/~djoyce/java/elements/bookVI/defVI3.html.

[7] Takuya Okabe. Biophysical optimality of the golden angle in phyllotaxis. *Scientific Reports*, 5 (1):15358, 2015. doi: 10.1038/srep15358.

[8] Christopher Bartlett. Nautilus spirals and the meta-golden ratio chi. *Nexus Network Journal*, 21(3):641–656, 2019. doi: 10.1007/s00004-018-0419-3.

[9] The Editors of Encyclopaedia Britannica. Parthenon. Encyclopaedia Britannica, n.d. URL https://www.britannica.com/topic/Parthenon.

[10] Faculty of Mathematics, University of Cambridge. Prof Hannah Fry. University of Cambridge, n.d. URL https://www.maths.cam.ac.uk/person/hf418.

[11] Hannah Fry. The (real) beauty of the Parthenon. YouTube Short, 2025. URL https://www.youtube.com/shorts/Y4tSre9bqQM.

[12] Farhad B. Naini. The golden ratio—dispelling the myth. *Maxillofacial Plastic and Reconstructive Surgery*, 46(1):2, 2024. doi: 10.1186/s40902-024-00411-2.

[13] National Eye Institute. Diabetic retinopathy. Web page, September 2025. URL https://www.nei.nih.gov/eye-health-information/eye-conditions-and-diseases/diabetic-retinopathy. Updated 11 September 2025.

[14] Piotr Kanclerz, Raimo Tuuminen, and Ramin Khoramnia. Imaging modalities employed in diabetic retinopathy screening: A review and meta-analysis. *Diagnostics*, 11(10):1802, 2021. doi: 10.3390/diagnostics11101802.

[15] Kai Jin, Xingru Huang, Jingxing Zhou, Yunxiang Li, Yan Yan, Yibao Sun, Qianni Zhang, Yaqi Wang, and Juan Ye. FIVES: A fundus image dataset for artificial intelligence based vessel segmentation. *Scientific Data*, 9(1):475, 2022. doi: 10.1038/s41597-022-01564-3.

[16] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2): 255–308, 2009. doi: 10.1090/S0273-0979-09-01249-X.

[17] RJ Nieto. Computational stability of cubical homology: Orthogonal controls for fault isolation in retinal diagnostics. GitHub repository, 2026. URL https://github.com/reyjosephnieto/ph. MIT License.

[18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016. ISBN 9780262035613. URL https://www.deeplearningbook.org/.

[19] Nathan Ballantyne. Epistemic trespassing. *Mind*, 128(510):367–395, 2019. doi: 10.1093/mind/fzx042.

[20] Texas A&M Institute of Data Science. What is data science? Texas A&M University, n.d. URL https://tamids.tamu.edu/about/what-is-data-science/.

[21] Australian Mathematical Sciences Institute. The language of maths. YouTube video, March 2019. URL https://www.youtube.com/watch?v=YtERoY3nVl8. Published 14 March 2019.

[22] Heinz-Dieter Ebbinghaus, Jörg Flum, and Wolfgang Thomas. *Mathematical Logic*. Springer, New York, 2 edition, 1994. doi: 10.1007/978-1-4757-2355-7.

[23] Harry Deutsch, Oliver Marshall, and Andrew David Irvine. Russell's Paradox. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2026 edition, 2026. URL https://plato.stanford.edu/archives/spr2026/entries/russell-paradox/. Substantively revised 13 March 2026.

[24] Institute for Molecular Bioscience, The University of Queensland. DNA, RNA, genes and genomics—what's the difference? Web page, n.d. URL https://imb.uq.edu.au/dna-rna-genes-and-genomics-%E2%80%93-whats-difference. Accessed 1 August 2026.

[25] Robert H. Whittaker. New concepts of kingdoms of organisms. *Science*, 163(3863):150–160, 1969. doi: 10.1126/science.163.3863.150.

[26] Leigh Zerboni, Nandini Sen, Stefan L. Oliver, and Ann M. Arvin. Molecular mechanisms of varicella zoster virus pathogenesis. *Nature Reviews Microbiology*, 12(3):197–210, 2014. doi: 10.1038/nrmicro3215. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4066823/.

[27] Spanning Tree. An introduction to propositional logic. YouTube video, December 2022. URL https://www.youtube.com/watch?v=5NGKbiA04Cw. Published 11 December 2022.