AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

Senior Secondary Australian Curriculum

General Mathematics Glossary

Financial Mathematics

Annuity

An **annuity** is a compound interest investment from which payments are made on a regular basis for a fixed period of time. At the end of this time the investment has no residual value.

Book value

The **book value** is the value of an asset recorded on a balance sheet. The book value is based on the original cost of the asset less depreciation.

For example, if the original cost of a printer is $500 and its value depreciates by $100 over the next year, then its book value at the end of the year is $400.

There are three commonly used methods for calculating yearly depreciation in the value of an asset, namely, **reducing balance depreciation**, **flat rate depreciation** or **unit cost depreciation**.

CPI

The **Consumer Price Index** (CPI) is a measure of changes, over time, in retail prices of a constant basket of goods and services representative of consumption expenditure by resident households in Australian metropolitan areas.

Effective annual rate of interest

The **effective annual rate of interest** $i_{\text{effective}}$ is used to compare the interest paid on loans (or investments) with the same nominal annual interest rate i but with different compounding periods (daily, monthly, quarterly, annually, other)

If the number of compounding periods per annum is n, then $i_{\text{effective}} = (1+\frac{i}{n})^n - 1$

For example if the quoted annual interest rate for a loan is 9%, but interest is charged monthly, then the effective annual interest rate charged is $i_{\text{effective}} = \left(1+\frac{0.09}{12}\right)^{12} - 1 = 0.9416...$, or around 9.4%.

Diminishing value depreciation see Reducing balance depreciation

Flat rate depreciation

In flat rate or straight-line depreciation the value of an asset is depreciated by a fixed amount each year. Usually this amount is specified as a fixed percentage of the original cost.

GST

The **GST** (Goods and Services Tax) is a broad sales tax of 10% on most goods and services transactions in Australia.

Straight-line depreciation

See: flat rate depreciation

Compound interest

The interest earned by investing a sum of money (the principal) is compound interest if each successive interest payment is added to the principal for the purpose of calculating the next interest payment.

For example, if the principal *P* earns compound interest at the rate of *i* % per period, then after *n* periods the total amount accrued is $P(1+\frac{i}{100})^n$. When plotted on a graph, the total amount accrued is seen to grow exponentially.

Perpetuity

A **perpetuity** is a compound interest investment from which payments are made on a regular basis in perpetuity (forever). This is possible because the payments made at the end of each period exactly equal the interest earned during that period.

Price to earnings ratio (of a share)

The price to earnings ratio of a share (P/E ratio) is defined as :

$$P/E\ ratio = \frac{Market\ price\ per\ share}{Annual\ earnings\ per\ share}$$

Reducing balance depreciation

In **reducing balance depreciation** the value of an asset is depreciated by a fixed percentage of its value each year.

Reducing balance depreciation is sometimes called diminishing value depreciation.

Reducing balance loan

A reducing balance loan is a compound interest loan where the loan is repaid by making regular payments and the interest paid is calculated on the amount still owing (the reducing balance of the loan) after each payment is made.

Simple interest

Simple interest is the interest accumulated when the interest payment in each period is a fixed fraction of the principal. For example, if the principle *P* earns simple interest at the rate of *i* % per period, then after *n* periods the accumulated simple interest is $nP\frac{i}{100}$

When plotted on a graph, the total amount accrued is seen to grow linearly.

Unit cost depreciation

In unit cost depreciation, the value of an asset is depreciated by an amount related to the number of units produced by the asset during the year.

Geometry and trigonometry

Angle of elevation

The angle a line makes above a plane.

Angle of depression

The angle a line makes below a plane.

Area of a triangle

The general rule for determining the area of a triangle is: $area = \frac{1}{2} base \times height$

Bearings (compass and true)

A **bearing** is the direction of a fixed point, or the path of an object, from the point of observation.

Compass bearings are specified as angles either side of north or south. For example a compass bearing of N50°E is found by facing north and moving through an angle of 50° to the East.

True (or three figure) bearings are measured in degrees from the north line. Three figures are used to specify the direction. Thus the direction of north is specified as 000°, east is specified as 090°, south is specified as 180° and north-west is specified as 315°.

Cosine rule

For a triangle of side lengths *a*, *b* and *c* and angles *A*, *B* and *C*, the **cosine rule** states that

$c^2 = a^2 + b^2 - 2ac \cos C$

Heron's rule

Heron's rule is a rule for determining the area of a triangle given the lengths of its sides.

The area *A* of a triangle of side lengths *a*, *b* and *c* is given by $A = \sqrt{s(s-a)(s-b)(s-c)}$ where $s = \frac{1}{2}(a+b+c)$.

Similar figures

Two geometric figures are similar if they are of the same shape but not necessarily of the same size.

Sine rule

For a triangle of side lengths *a*, *b* and *c* and angles *A*, *B* and *C*, the **sine rule** states that

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

Triangulation

The process of determining the location of a point by measuring angles to it from known points at either end of a fixed baseline, rather than measuring distances to the point directly. The point can then be fixed as the third point of a triangle with one known side and two known angles.

Scale factor

A **scale factor** is a number that scales, or multiplies, some quantity. In the equation $y = kx$, k is the scale factor for x.

If two or more figures are similar, their sizes can be compared. The scale factor is the ratio of the length of one side on one figure to the length of the corresponding side on the other figure. It is a measure of magnification, the change of size.

Graphs & networks

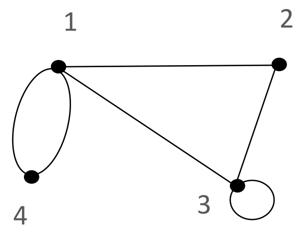
Adjacent (graph) see graph

Adjacency matrix

An **adjacency matrix** for a non-directed graph with n vertices is a $n \times n$ matrix in which the entry in row i and column j is the number of edges joining the vertices i and j. In an adjacency matrix, a **loop** is counted as 1 edge.

Example:

Non-directed graph



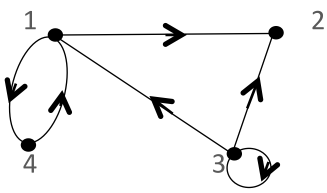
Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

For a directed graph the entry in row i and column j is the number of directed edges (arcs) joining the vertex i and j in the direction i to j.

Example:

Directed graph



Adjacency matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Algorithm

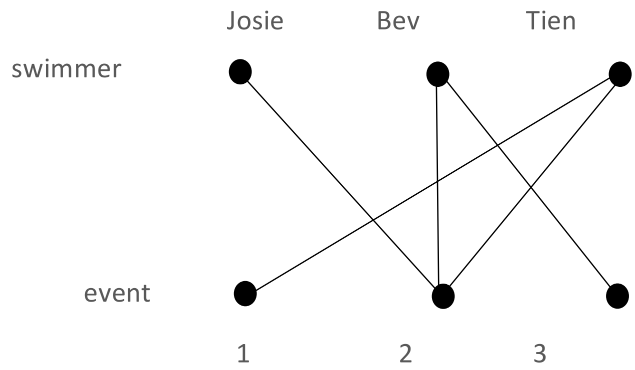
An **algorithm** is a precisely defined routine procedure that can be applied and systematically followed through to a conclusion. An example is **Prim's algorithm** for determining a **minimum spanning tree** in a network.

Arc **see directed graph**

Bipartite Graph

A bipartite graph is a graph whose set of vertices can be split into two distinct groups in such a way that each edge of the graph joins a vertex in the first group to a vertex in the second group.

Example:



Bridge see connected graph

Closed path

See path

Closed trail

See trail

Closed walk

See walk

Complete graph

A **complete graph** is a **simple graph** in which every vertex is joined to every other vertex by an edge.

The complete graph with n vertices is denoted K_n .

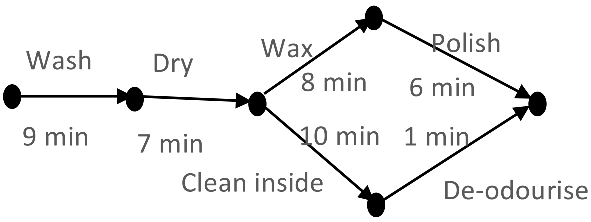
Connected graph

A graph is **connected** if there is a path between each pair of vertices. A **bridge** is an edge in a connected graph that, if removed, leaves a graph disconnected.

Critical path analysis (CPA)

A project often involves many related activities some of which cannot be started until one or more earlier tasks have been completed. One way of scheduling such activities that takes this into account is to construct a network diagram.

The network diagram below can be used to schedule the activities of two or more individuals involved in cleaning and polishing a car. The completion times for each activity are also shown.



Critical path analysis is a method for determining the longest path (the **critical path**) in such a network and hence the minimum time in which the project can be completed. There may be more that one critical path in the network. In this project the critical path is 'Wash-Dry-Wax-Polish' with a total completion time of 30 minutes.

The **earliest starting time (EST)** of an activity 'Polish' is 24 minutes because activities 'Wash', 'Dry' and 'Wax' must be completed first. The process of systematically determining earliest starting times is called **forward scanning.**

The shortest time that the project can be completed is 30 minutes. Thus, the **latest starting time (LST)** for the activity 'De-odourise' is 29 minutes. The process of systematically determining latest starting times is called **backward scanning**.

Float or slack

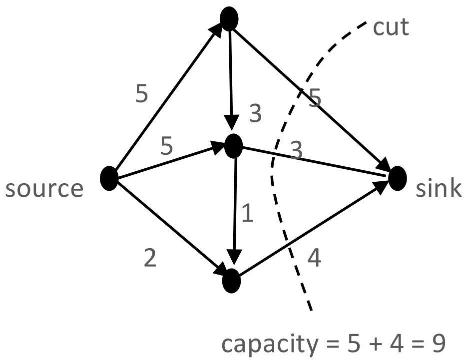
Is the amount of time that a task in a project network can be delayed without causing a delay to subsequent tasks. For example, the activity 'De-odourise' is said to have a **float** of 3 minutes because its earliest EST (26 minutes) is three minutes before its LST (29 minutes). As a result this activity can be started at any time between 26 and 29 minutes after the project started. All activities on a critical path have zero floats.

Cut (in a flow network)

In a flow network, a **cut** is a partition of the vertices of a graph into two separate groups with the **source** in one group and the **sink** in the other.

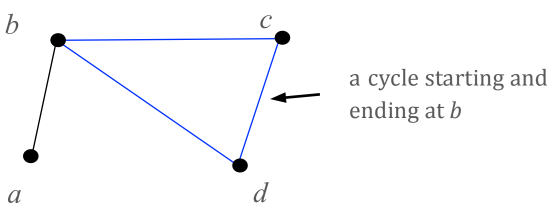
The capacity of the **cut** is the sum of the capacities of the cut edges directed from source to sink. Cut edges directed from sink to source are ignored.

Example:



Cycle

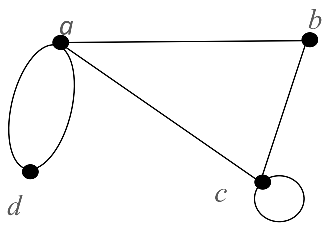
A **cycle** is a closed walk begins and starts at the same vertex and in which has no repeated edges or vertices except the first. If *a*, *b*, *c* and *d* are the vertices of a graph, the closed walk *bcdb* that starts and ends at vertex *b* (shown in blue) an example of a cycle.



Degree of a vertex (graph)

In a graph, the **degree of a vertex** is the number of edges incident with the vertex, with loops counted twice. It is denoted deg *v.*

In the graph below, deg *a* = 4, deg *b* = 2, deg *c* = 4 and deg *d*=2.

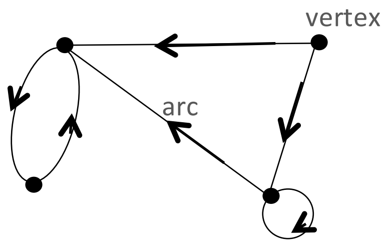


Digraph

See directed graph

Directed graph

A **directed graph** is a diagram comprising points, called vertices, joined by directed lines called **arcs**. The directed graphs are commonly called **digraphs**.



Earliest starting time (EST)

See Critical Path Analysis

Edge

See graph

Euler's formula

For a connected planar graph, **Euler's rule** states that

$$v + f - e = 2$$

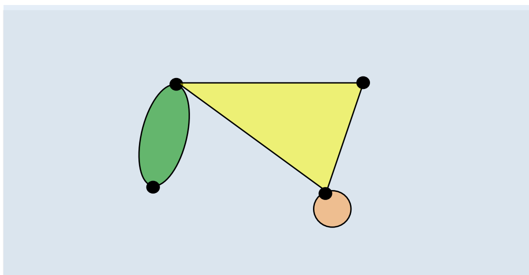
where v is the number vertices, e the number of edges and f is the number of faces.

Eulerian graph

A connected graph is **Eularian** if it has a closed trail (starts and ends at the same vertex), that is, includes every edge and once only; such a trail is called an **Eulerian trail**. An Eularian trial may include repeated vertices. A connected graph is **semi-Eularian** if there is an open trail that includes every once only.

Face

The faces of a planar graph are the regions bounded by the edges including the outer infinitely large region. The planar graph shown has four faces.



Float time

See Critical Path Analysis

Flow network

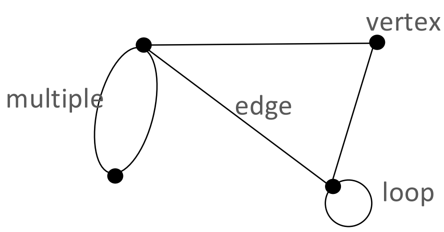
A **flow network** is a directed graph where each edge has a capacity (e.g. 100 cars per hour, 800 litres per minute, etc) and each edge receives a flow. The amount of flow on an edge cannot exceed the capacity of the edge. A flow must satisfy the restriction that the amount of flow into a node equals the amount of flow out of it, except when it is a **source**, which has more outgoing flow, or a **sink**, which has more incoming flow. A flow network can be used to model traffic in a road system, fluids in pipes, currents in an electrical circuit, or any situation in which something travels through a network of nodes.

Food web

A **food web** (or food chain) depicts feeding connections (who eats whom) in an ecological community.

Graph

A **graph** is a diagram that consists of a set of points, called **vertices** that are joined by a set of lines called **edges**. Each edge joins two vertices. A **loop** is an edge in a graph that joins a **vertex** in a **graph** to itself. Two vertices are **adjacent** if they a joined by an edge. Two or more edges connect the same vertices are called **multiple edges**.



Hamiltonian cycle

A **Hamiltonian cycle** is a **cycle** that includes each **vertex** in a **graph (except the first),** once only.

Hamiltonian path

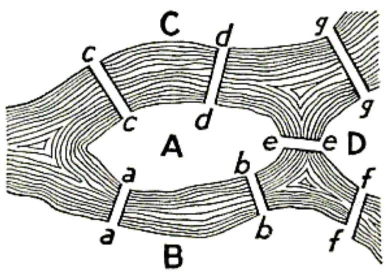
A **Hamilton path** is path that includes every **vertex** in a **graph** once only. A Hamilton path that begins and ends at the same vertex is a Hamiltonian cycle.

Hungarian algorithm

The Hungarian algorithm is used to solve assignment (allocation) problems.

Königsberg bridge problem

The Königsberg bridge problem asks: Can the seven bridges of the city of Königsberg all be traversed in a single trip that starts and finishes at the same place?



ACARA AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

Latest starting time (LST)

See Critical Path Analysis

Length (of a walk)

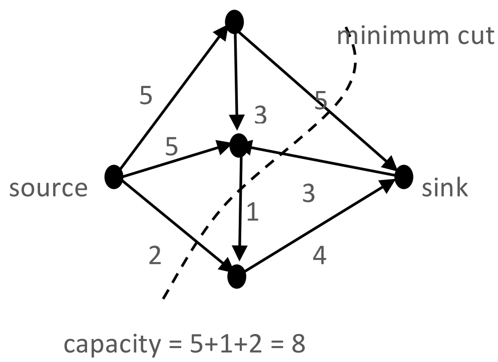
The **length** of a walk is the number of edges it includes.

Minimum cut-maximum flow theorem

The **maximum flow–minimum cut theorem** states that in a flow network, the maximum flow from the **source** to the **sink** is equal to the capacity of the **minimum cut**.

In everyday language, the minimum cut involves identifies the 'bottle-neck' in the system.

Example:



Minimum spanning tree

For a given connected **weighted graph**, the **minimum spanning tree** is the **spanning tree** of minimum length.

Multiple edges see graph

Network

The word network is frequently used in everyday life, e.g. television network, rail network, etc. Weighted graphs or digraphs can often be used to model such networks.

Open path

See path

Open walk

See walk

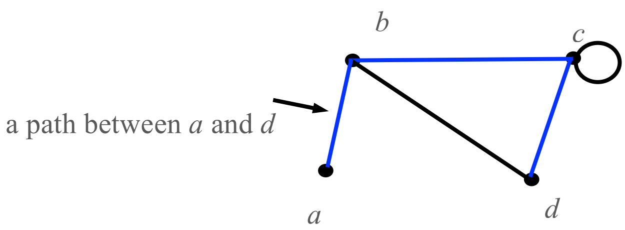
Open trail

See trail

Path (in a graph)

A **path** in a graph is a walk in which all of the edges and all the vertices are different.. A path that starts and finishes at different vertices is said to be open, while a path that stats and finishes at the same vertex is said to be closed. A cycle is a closed path.

If *a* and *d* are the vertices of a graph, a walk from *a* to *d* along the edges coloured blue is a path. Depending on the graph, there may be multiple paths between the same two vertices, as is the case here.



Planar graph

A **planar graph** is a graph that can be drawn in the plane. A planar graph can always be drawn so that no two edges cross.

Prim's algorithm

An **algorithm** for determining a **minimum spanning tree** in a connected weighted graph.

Round-robin sporting competition

A single round robin sporting competition is a competition in which each competitor plays each other competitor once only.

Semi-Eularian graph

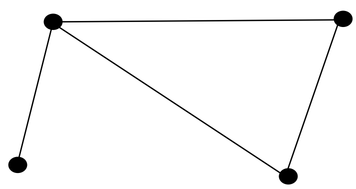
See Eularian graph

Semi-Hamiltonian graph

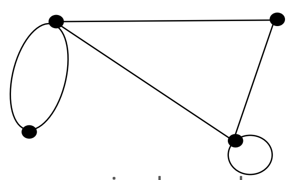
See Hamiltonian graph

Simple graph

A simple graph has no loops or multiple edges.



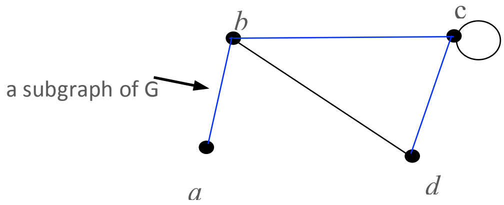
simple graph



non-simple graph

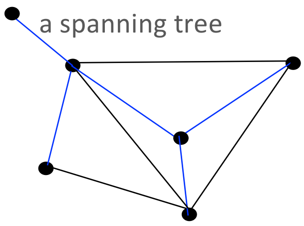
Subgraph

When the vertices and edges of a graph *A* (shown in blue) are the vertices and edges of the graph *G*, graph *A* is said to be a **subgraph** of graph *G*.



Spanning tree

A **spanning tree** is a **subgraph** of a **connected graph** that connects all vertices and is also a **tree.**



Trail

A **trail** is a **walk** in which no edge is repeated.

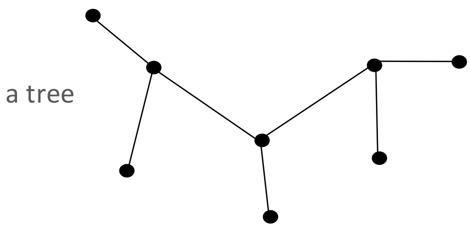
The travelling salesman problem

The travelling salesman problem can be described as follows: Given a list of cities and the distance between each city, find the shortest possible route that visits each city exactly once.

While in simple cases this problem can be solved by systematic identification and testing of possible solutions, no there is no known efficient method for solving this problem.

Tree

A **tree** is a connected graph with no cycles.



Vertex

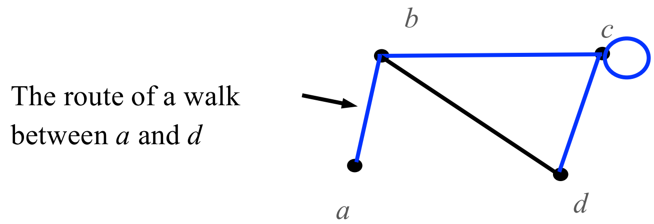
See graph

Walk (in a graph)

A **walk** in a graph is a sequence of vertices such that from each of its vertices there is an edge to the next vertex in the sequence. A walk that starts and finishes at different vertices is said to be an **open walk**. A walk that starts and finishes at the same vertex is said to be c**losed walk**.

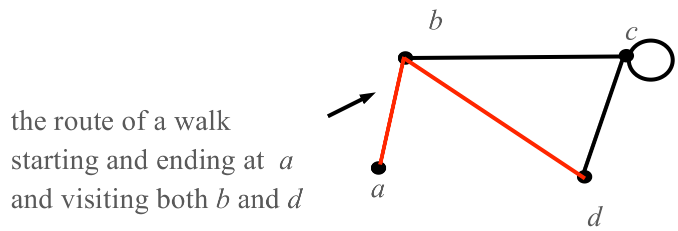
If *a*, *b*, *c* and *d* are the vertices of a graph with edges *ab*, *bc*, *cc*, *cd* and *bd*, then the sequence of edges (*ab*, *bc*, *cc*, *cd*) constitute a walk. The route followed on this walk is shown in blue on the graph below.

This walk is denoted by the sequence of vertices *abccd*. The walk is open because it begins and finishes at different vertices.



A walk can include repeated vertices (as is the case above) or repeated edges.

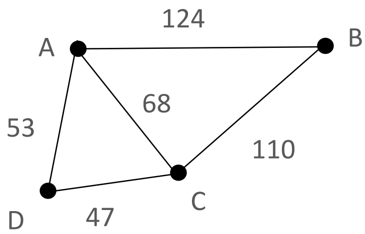
A example of a closed walk with both repeated edges and hence vertices is defined by the sequence of edges (*ab*, *bd*, *db*, *ba*) and is denoted by the sequence of vertices *abdba*. The route followed is shown in red in the graph below.



Depending on the graph, there may be multiple walks between the same two vertices, as is the case here.

Weighted graph

A weighted graph is a graph in which each edge is labelled with a number used to represent some quantity associated with the edge. For example, if the vertices represent towns, the weights on the edges may represent the distances in kilometres between the towns.



Growth and decay in sequences

Arithmetic sequence

An **arithmetic sequence** is a sequence of numbers such that the difference between any two successive members of the sequence is constant.

For example, the sequence

2, 5, 8, 11, 14, 17, …

is an arithmetic sequence with first term 2 and common difference 3.

By inspection of the sequence, the rule for the *n*th term t_n of this sequence is:

$t_n = 2 + (n - 1)3 = 3n - 1 \quad n \geq 1$

If t_n is used to denote the *n*th term in the sequence, then a recursion relation that will generate this sequence is: $t_1 = 2$, $t_{n+1} = t_n + 3 \quad n \geq 1$

Break-even point

The **break-even point** is the point at which revenue begins to exceed the cost of production.

First-order linear recurrence relation

A **first-order linear recurrence relation** is defined by the rule: $t_0 = a$, $t_{n+1} = bt_n + c$ for $n \geq 1$

For example, the rule: $t_0 = 10$, $t_n = 5t_{n-1} + 1$ for $n \geq 1$ is a first-order recurrence relation.

The sequence generated by this rule is: 10, 51, 256, … as shown below.

$t_1 = 10$, $t_2 = 5t_1 + 1 = 5 \times 10 + 1 = 51$, $t_3 = 5t_2 + 1 = 5 \times 51 + 1 = 256$, …

Geometric growth or decay (sequence)

A sequence displays geometric growth or decay when each term is some constant multiple (greater or less than one) of the preceding term. A multiple greater than one corresponds to growth. A multiple less than one corresponds to decay.

For example, the sequence:

> 1, 2, 4, … displays geometric growth because each term is double the previous term.
>
> 100, 10, 0.1, … displays geometric decay because each term is one tenth of the previous term.

Geometric growth is an example of exponential growth in discrete situations.

Geometric sequence

A **geometric sequence**, is a sequence of numbers where each term after the first is found by multiplying the previous one by a fixed non-zero number called the **common ratio**. For example, the sequence

> 2, 6, 18, ...

is a geometric sequence with first term 2 and common ratio 3.

By inspection of the sequence, the rule for the *n*th term of this sequence is:

$$t_n = 2 \times 3^{n-1} \quad n \geq 1$$

If t_n is used to denote the *n*th term in the sequence, then a recursion relation that will generate this sequence is: $t_1 = 2$, $t_{n+1} = 3t_n \quad n \geq 1$

Linear growth or decay (sequence)

A sequence displays linear growth or decay when the difference between successive terms is constant. A positive constant difference corresponds to linear growth while a negative constant difference corresponds to decay.

Examples:

The sequence, 1, 4, 7, … displays linear growth because the difference between successive terms is 3.

The sequence, 100, 90, 80, … displays linear decay because the difference between successive terms is –10. By definition, arithmetic sequences display linear growth or decay.

Recursion

See recurrence relation

Recurrence relation

A **recurrence relation** is an equation that recursively defines a sequence; that is, once one or more initial terms are given, each further term of the sequence is defined as a function of the preceding terms.

Sequence

A **sequence** is an ordered list of numbers (or objects).

For example 1, 3, 5, 7 is a sequence of numbers that differs from the sequence 3, 1, 7, 5 as order matters.

A sequence maybe finite, for example, 1, 3, 5, 7 (the sequence of the first four odd numbers), or infinite, for example, 1, 3, 5, … (the sequence of all odd numbers).

Linear equations (relations) and graphs

Linear equation

A linear equation in one variable x is an equation of the form $ax + b = 0$, e.g. $3x + 1 = 0$

A linear equation in two variables x and y is an equation of the form $ax + by + c = 0$,

e.g. $2x – 3y + 5 = 0$

Linear graph

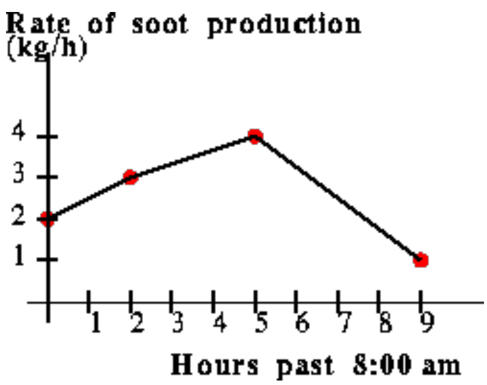
A **linear graph** is a graph of a linear equation with two variables. If the linear equation is written in the form $y = a + bx$, then a represents the y-intercept and b represents the slope (or gradient) of the linear graph.

Piecewise-linear graph

A graph consisting of one or more none overlapping line segments.

Sometimes called a line segment graph.

Example:



Slope (gradient)

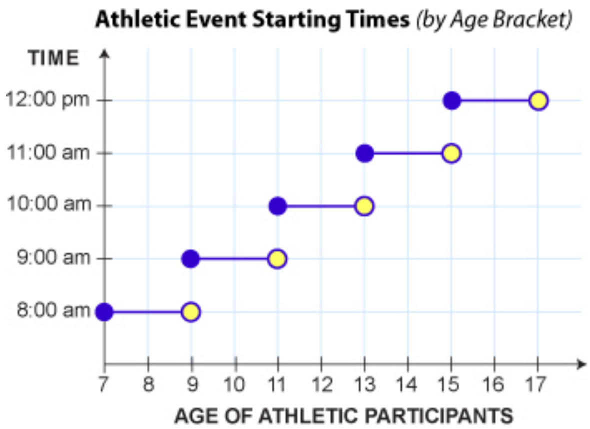
The **slope** or **gradient** of a line describes its steepness, incline, or grade.

Slope is normally described by the ratio of the "rise" divided by the "run" between two points on a line.

See also linear graph.

Step graph

A graph consisting of one or more non-overlapping horizontal line segments that follow a step-like pattern.



Matrices

Addition of matrices

If ***A*** and ***B*** are matrices of the same size (order) and the elements of **A** are a_{ij} and the elements of **B** are b_{ij} then the elements of ***A* + *B*** are $a_{ij} + b_{ij}$

For example if $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & 4 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 5 & 1 \\ 2 & 1 \\ 1 & 6 \end{bmatrix}$ then

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \\ 2 & 10 \end{bmatrix}$$

Elements (Entries) of a matrix

The symbol a_{ij} represents the (i,j) element occurring in the i^{th} row and the j^{th} column.

For example a general 3 × 2 matrix is:

$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$ where a_{32} is the element in the third row and the second column

Identity matrix

A multiplicative **identity matrix** is a square matrix in which all of the elements in the leading diagonal are 1s and the remaining elements are 0s. Identity matrices are designated by the letter ***I***.

For example,

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ are both identity matrices.

.

There is an identity matrix for each size (or order) of a square matrix. When clarity is needed, the order is written with a subscript: $\boldsymbol{I_n}$

Inverse of a square matrix

The **inverse of a square matrix *A*** is written as $\boldsymbol{A}^{-1}$ and has the property that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

Not all square matrices have an inverse. A matrix that has an inverse is said to be **invertible**.

Inverse of a 2 × 2 matrix

The **inverse** of the matrix $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ Is

$\boldsymbol{A}^{-1} = \dfrac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ provided $ad - bc \neq 0$

Leading diagonal

The leading diagonal of a square matrix is the diagonal that runs from the top left corner to the bottom right corner of the matrix.

Matrix (matrices)

A **matrix** is a rectangular array of elements or entities displayed in rows and columns.

For example,

$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & 4 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 1 & 8 & 0 \\ 2 & 5 & 7 \end{bmatrix}$ are both matrices with six elements.

Matrix ***A*** is said to be a 3×2 matrix (three rows and two columns) while ***B*** is said to be a 2 × 3 matrix (two rows and three columns).

A **square matrix** has the same number of rows and columns.

A **column matrix** (or vector) has only one column.

A **row matrix** (or vector) has only one row.

Matrix multiplication

Matrix multiplication is the process of multiplying a matrix by another matrix.

For example, forming the product

$$\begin{bmatrix} 1 & 8 & 0 \\ 2 & 5 & 7 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 25 \\ 11 & 45 \end{bmatrix}$$

The multiplication is defined by $1 \times 2 + 8 \times 0 + 0 \times 4 = 2$

$1 \times 1 + 8 \times 3 + 0 \times 4 = 25$

$2 \times 2 + 5 \times 0 + 7 \times 1 = 11$

$2 \times 1 + 5 \times 3 + 7 \times 4 = 45$

This is an example of the process of matrix multiplication.

The product ***AB*** of two matrices ***A*** and ***B*** of size $m \times n$ and $p \times q$ respectively is defined if $n = p$.

If $n = p$ the resulting matrix has size $m \times q$.

If $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$ then

$$\boldsymbol{AB} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} \end{bmatrix}$$

Order (of a matrix)

See **size** (of a matrix)

Scalar multiplication (matrices)

Scalar multiplication is the process of multiplying a matrix by a scalar (number).

For example, forming the product

$$10 \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 20 & 10 \\ 0 & 30 \\ 10 & 40 \end{bmatrix}$$

is an example of the process of scalar multiplication.

In general for the matrix ***A*** with elements a_{ij} the elements of $k\boldsymbol{A}$ are ka_{ij}.

Singular matrix

A matrix is singular if det ***A*** = 0. A singular matrix does not have a multiplicative inverse.

Size (of a matrix)

Two matrices are said to have the same **size** (or **order**) **if** they have the same number of rows and columns. A matrix with *m* rows and *n* columns is said to be a $m \times n$ matrix.

For example, the matrices

$$\begin{bmatrix} 1 & 8 & 0 \\ 2 & 5 & 7 \end{bmatrix} \text{ and } \begin{bmatrix} 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$$

have the same size. They are both 2×3 matrices.

Zero matrix

A zero matrix is a matrix if all of its entries are zero. For example:

$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ are zero matrices.

Statistics

Association

A general term used to describe the relationship between two (or more) variables. The term **association** is often used interchangeably with the term **correlation**. The latter tends to be used when referring to the strength of a linear relationship between two numerical variables.

Average percentage method

In the **average percentage method** for calculating a **seasonal index**, the data for each 'season' are expressed as percentages of the average for the year. The percentages for the corresponding 'seasons' for different years are then averaged using a mean or median to arrive at a seasonal index.

Categorical variable

A **categorical variable** is a variable whose values are categories.

Examples include blood group (A, B, AB or O) or house construction type (brick, concrete, timber, steel, other).

Categories may have numerical labels, eg. the numbers worn by player in a sporting team, but these labels have no numerical significance, they merely serve as labels.

Categorical data

Data associated with a **categorical variable** is called categorical data.

Causation

A relationship between an explanatory and a response variable is said to be causal if the change in the explanatory variable actually causes a change in the response variable. Simply knowing that two variables are associated, no matter how strongly, is not sufficient evidence by itself to conclude that the two variables are causally related.

Possible explanations for an observed association between an explanatory and a response variable include:

- the **explanatory variable** is actually causing a change in the response variable
- there may be causation, but the change may also be caused by one or more uncontrolled variables whose effects cannot be disentangled from the effect of the response variable. This is known as **confounding**.
- there is no causation, the association is explained by at least one other variable that is associated with both the explanatory and the response variable . This is known as a **common response**.
- the **response variable** is actually causing a change in the explanatory variable

Coefficient of determination

In a linear model between two variables, the coefficient of determination (R^2) is the proportion of the total variation that can be explained by the linear relationship existing between the two variables, usually expressed as a percentage. For two variables only, the coefficient of determination is numerically equal to the square of the correlation coefficient (r^2).

Example

A study finds that the correlation between the heart weight and body weight of a sample of mice is $r = 0.765$. The coefficient of determination = $r^2 = 0.765^2 = 0.5852$ … or approximately 59%

From this information, it can be concluded that approximately 59% of the variation in heart weights of these mice can be explained by the variation in their body weights.

Note: The coefficient of determination has a more general and more important meaning in considering relationships between more than two variables, but this is not a school level topic.

Common response
See Causation

Confounding

See Causation

Continuous data

Data associated with a **continuous variable** is called continuous data.

Continuous variable

A **continuous variable** is a **numerical variable** that can take any value that lies within an interval. In practice, the values taken are subject to accuracy of the measurement instrument used to obtain these values.

Examples include height, reaction time and systolic blood pressure.

Correlation

Correlation is a measure of the strength of the linear relationship between two variables. See also **association**.

Correlation coefficient (*r*)

The correlation coefficient (*r*) is a measure of the strength of the liner relationship between a pair of variables. The formula for calculating *r* is given below.

For variables *x* and *y*, and computed for *n* cases, the formula for *r* is:

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

Discrete data

Discrete data is data associated with a discrete variable. Discrete data is sometimes called count data.

Discrete variable

A **discrete variable** is a **numerical variable** that can take only integer values.

Examples include the number of people in a car, the number of decayed teeth in 18 year-old males, etc.

Explanatory variable

When investigating relationships in bivariate data, the **explanatory variable** is the variable used to explain or predict a difference in the **response variable**.

For example, when investigating the relationship between the temperature of a loaf of bread and the time it has spent in a hot oven, *temperature* is the response variable and *time* is the explanatory variable.

Extrapolation

In the context of fitting a linear relationship between two variables, extrapolation occurs when the fitted model is used to make predictions using values of the **explanatory variable** that are outside the range of the original data. Extrapolation is a dangerous process as it can sometimes lead to quite erroneous predictions.

See also interpolation.

Five-number summary

A **five-number summary** is a method of summarising a set of data using the minimum value, the lower or first-quartile (Q_1), the median, the upper or third-quartile (Q_3) and the maximum value. Forms the basis for a boxplot.

Interpolation

In the context of fitting a linear relationship between two variables, interpolation occurs when the fitted model is used to make predictions using values of the **explanatory variable** that lie within the range of the original data.

See also extrapolation.

Irregular variation or noise (time series)

Irregular variation or noise is erratic and short-term variation in a time series that is the product of chance occurrences.

Least-squares line

In fitting a straight-line $y = a + bx$ to the relationship between a response variable y and an explanatory variable x, the **least-squares line** is the line for which the sum of the squared **residuals** is the smallest.

The formula for calculating the slope (b) and the intercept (a) of the least squares line is given below.

For variables x and y computed for n cases, the slope (b) and intercept (a) of the least-squares line are given by:

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \text{ or } b = r\frac{s_y}{s_x} \qquad \text{and} \qquad a = \bar{y} - b\bar{x}$$

Location

The notion of central or 'typical value' in a sample distribution.

See also **mean**, **median** and **mode**.

Mean

The arithmetic mean of a list of numbers is the sum of the data values divided by the number of values in the list.

In everyday language, the arithmetic mean is commonly called the average.

For example, for the following list of five numbers 2, 3, 3, 6, 8 the mean equals

$$\frac{2+3+3+6+8}{5} = \frac{22}{5} = 4.4$$

In more general language, the mean of n observations $x_1, x_2, \ldots, x_n$ is $\bar{x} = \frac{\sum x_i}{n}$

Median

The **median** is the value in a set of ordered set of data values that divides the data into two parts of equal size. When there are an odd number of data values, the median is the middle value. When there is an even number of data values, the median is the average of the two central values.

Mode

The **mode** is the most frequently occurring value is a data set.

Moving average

In a time series, a simple moving average is a method used to **smooth** the time series whereby each observation is replaced by a simple average of the observation and its near neighbours. This process reduces the effect of non-typical data and makes the overall trend easier to see.

Note: There are times when it is preferable to use a weighted average rather simple average, but this is not required in the current curriculum.

Outlier

An outlier in a set of data is an observation that appears to be inconsistent with the remainder of that set of data. An outlier is a surprising observation.

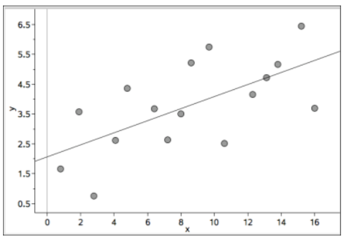
Residual values

The difference between the observed value and the value predicted by a statistical model (e.g., by a least-squares line)

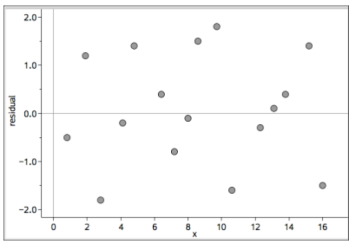
Residual plot

A residual plot is a **scatterplot** with the **residual values** shown on the vertical axis and the **explanatory variable** shown on the horizontal axis. Residual plots are useful in assessing the fit of the statistical model (e.g., by a least-squares line).

When the least-squares line captures the overall relationship between the response variable *y* and the explanatory variable *x*, the residual plot will have no clear pattern (be random) see opposite. This is what is hoped for.



scatterplot with least squares line



residual plot

If the least-squares line fails to capture the overall relationship between a response variable and an explanatory variable, a residual plot will reveal a pattern in the residuals. A residual plot will also reveal any outliers that may call into question the use of a least-squares line to describe the relationship. Interpreting patterns in residual plots is a skilled art and is not required in this curriculum.

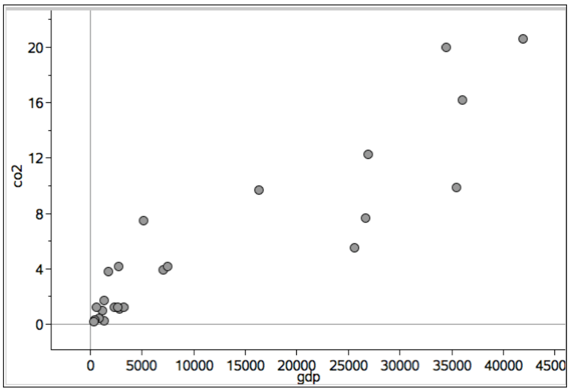
Response variable

See Explanatory variable

Scatterplot

A two-dimensional data plot using Cartesian co-ordinates to display the values of two variables in a bivariate data set.

For example the scatterplot below displays the CO_2 emissions in tonnes per person (*co2*) plotted against Gross Domestic Product per person in $US (*gdp*) for a sample of 24 countries in 2004. In constructing this scatterplot, gdp has been used as the explanatory variable.



Seasonal adjustment (adjusting for seasonality)

A term used to describe a time series from which periodic variations due to seasonal effects have been removed.

See also seasonal index.

Seasonal index

The seasonal index can be used to remove seasonality from data. An index value is attached to each period of the time series within a year. For the seasons of the year (Summer, Autumn, Winter, Spring) there are four separate seasonal indices; for months, there are 12 separate seasonal indices, one for each month, and so on. There are several methods for determining seasonal indices.

Seasonal variation

A regular rise and fall in the time series that recurs each year.

Seasonal variation is measured in terms of a **seasonal index**.

Smoothing (time series) **see moving average**

Standard deviation

The standard deviation is a measure of the variability or spread of a data set. It gives an indication of the degree to which the individual data values are spread around their **mean**.

The standard deviation of n observations $x_1, x_2, \ldots, x_n$ is

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Statistical investigation process

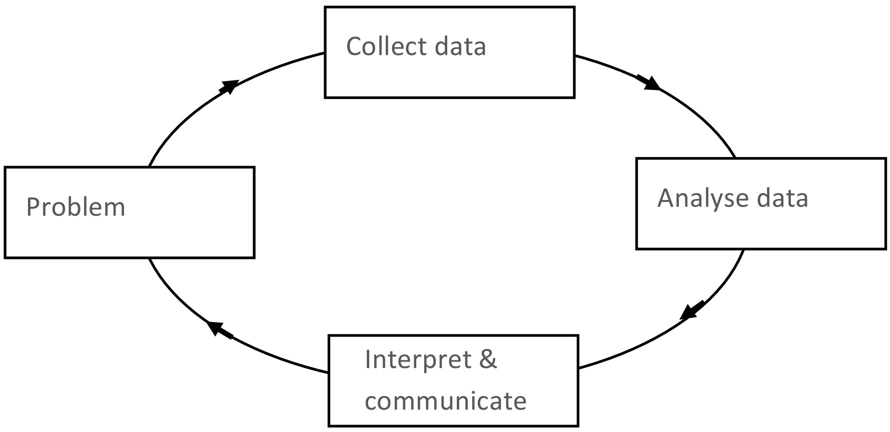
The statistical investigation process is a cyclical process that begins with the need to solve a real world problem and aims to reflect the way statisticians work. One description of the statistical investigation process in terms of four steps is as follows.

Step 1. Clarify the problem and formulate one or more questions that can be answered with data.

Step 2. Design and implement a plan to collect or obtain appropriate data.

Step 3. Select and apply appropriate graphical or numerical techniques to analyse the data.

Step 4. Interpret the results of this analysis and relate the interpretation to the original question; communicate findings in a systematic and concise manner.



Time series

Values of a variable recorded, usually at regular intervals, over a period of time. The observed movement and fluctuations of many such series comprise long-term **trend**, **seasonal variation**, and **irregular variation** or **noise**.

Time series plot

The graph of a **time series** with time plotted on the horizontal axis.

Trend (time series)

Trend is the term used to describe the general direction of a time series (increasing/ decreasing) over a long period of time.

Two-way frequency table

A two-way frequency table is commonly used for displaying the two-way **frequency distribution** that arises when a group of individuals or objects are categorised according to two criteria.

For example, the two-way table below displays the frequency distribution that arises when 27 children are categorised according to *hair type* (straight or curly) and *hair colour* (red, brown, blonde, black).

Hair colour	Hair type		Total
	Straight	Curly	
red	1	1	2
brown	8	4	12
blonde	1	3	4
black	7	2	9
Total	17	10	27

The row and column totals represent the total number of observations in each row and column and are sometimes called **row sums** or **column sums**.

If the table is 'percentaged' using row sums the resulting percentages are called **row percentages**. If the table is 'percentaged' using row sums the resulting percentages are called **column percentages**.