GAC010 Mathematics II: Probability, Statistics and Finance

GAC Mathematics

What this module is, and how it is marked

A cheaper headline rate need not produce the cheaper total, and a larger survey need not produce a more representative result. Mathematics II uses explicit rules, calculations and evidence limits to examine such questions.

GAC010 covers sequences, finance, probability and statistics. Your centre's current brief determines assessment tasks, weights, tools and deadlines. A **test**, **projects**, an **examination** and **coursework** may use different instructions; these original practice sheets do not establish their official pattern or university credit.

Show the known data, why the formula applies, the symbolic equation and the substitution. State the unit and interpret the result. A statistical summary describes a defined set of observations or a stated model; it does not automatically establish a cause or a whole-population finding.

All financial amounts, rates, fees and pay rules in the examples are invented teaching conditions. Use their stated timing and rounding assumptions rather than assuming an actual tax or lending rule.

Sequences and series

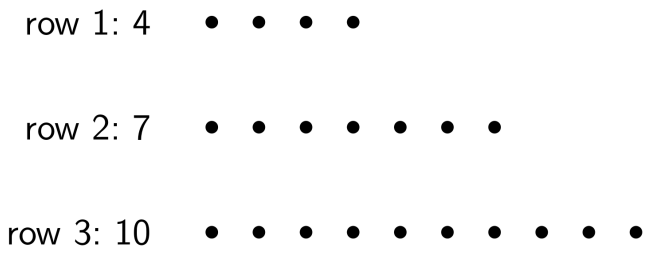
A **sequence** is an ordered list of numbers; a **series** is their sum.

An **arithmetic sequence** adds a fixed **common difference** d each step. A **geometric sequence** multiplies by a fixed **common ratio** r.

$$u_n = a + (n-1)d \qquad \text{and} \qquad u_n = ar^{n-1}$$

- The sums are $S_n = \frac{n}{2}(2a + (n-1)d)$ and $S_n = \dfrac{a(1-r^n)}{1-r}$ for $r \neq 1$.
- A geometric series with $|r| < 1$ has a **sum to infinity** $S_\infty = \dfrac{a}{1-r}$.
- Keep a term separate from a total. Some financial models compound geometrically, while others use simple interest or cash flows. Match the model to the specified rule.

> **Worked example.** row counts differ by a constant amount



Known: row counts are 4, 7 and 10, continuing with the same difference. Thus $a = 4$ and $d = 3$. For the fifth row:

$$u_n = a + (n-1)d$$

$$u_5 = 4 + (5-1)(3) = 16$$

For the first five rows together, use the arithmetic sum, not the fifth term.

$$S_n = \frac{n}{2}[2a + (n-1)d]$$

$$S_5 = \frac{5}{2}[2(4) + (5-1)(3)] = 50$$

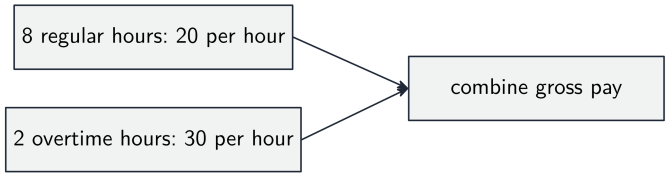
Practice sheet **2.1** gives a progressively harder set with worked solutions.

Finance: earning money and purchasing goods

This unit is the arithmetic of an ordinary life, in English.

- **Gross pay** is what you earn; **net pay** is what arrives after **deductions** such as tax and insurance.
- **Simple interest** is $I = Prt$ – interest on the original amount only.
- **Compound interest** is $A = P(1+r)^n$ – interest on the interest, which is why it is geometric.
- **Depreciation** is a reduction in recorded value. A fixed-percentage model uses $V_n = V_0(1-r)^n$; other depreciation conventions can use different rules.
- A **discount** of 20% multiplies by 0.8. Two successive discounts of 20% and 10% multiply by $0.8 \times 0.9 = 0.72$, which is a 28% reduction, not 30%.

> **Worked example.** calculate regular and overtime hours separately



Known: 8 hours at 20 units per hour and 2 overtime hours at 1.5 times that rate. Separate the two categories before adding pay.

$$G = h_r p + h_o k p$$

$$G = 8(20) + 2(1.5)(20) = 220 \text{ units}$$

If the invented fixed deduction is 15 units, take it from gross pay once.

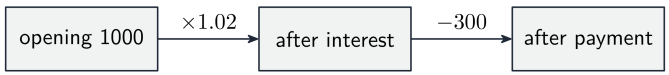
$$N = G - D = 220 - 15 = 205 \text{ units}$$

Practice sheet **2.2** gives a progressively harder set with worked solutions.

Finance: investing money and loans

- Under a stated constant discount rate and timing assumption, the **present value** of a future amount is: $PV = \dfrac{FV}{(1+r)^n}$.
- An **annuity** has specified payments at regular intervals. A **loan repayment** schedule may instead have changing amounts or an adjusted final payment; specify whether interest comes before or after payment.
- The label **APR** is an annual rate measure whose calculation and included charges depend on the stated convention. This sheet uses explicitly defined period rates, not an unstated APR convention.
- △ A nominal 12% teaching rate with monthly rate 1% has effective annual growth $(1.01)^{12} - 1 \approx 12.68\%$. A different compounding frequency or fee schedule needs a different calculation.

Worked example. interest then payment, in that order



Known: opening balance 1000 units, period rate 2%, end-of-period payment 300 units. Interest comes before payment under this stated convention.

$$B_1 = B_0(1+r) - M$$

$$B_1 = 1000(1.02) - 300 = 720$$

The next period's interest uses 720, not the original principal.

$$B_2 = B_1(1+r) - M = 720(1.02) - 300 = 434.40$$

Practice sheet **2.3** gives a progressively harder set with worked solutions.

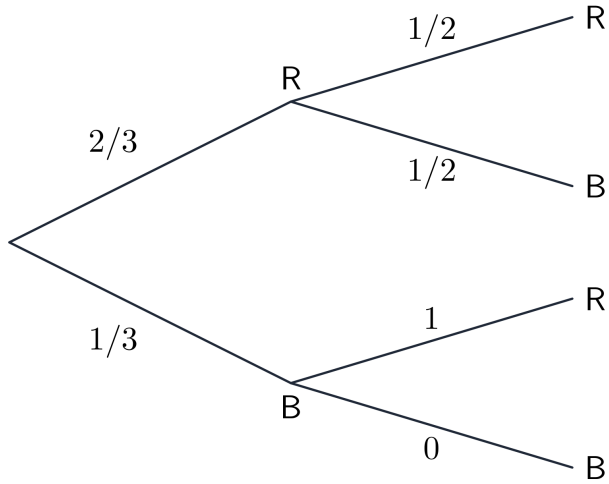
Probability: key concepts

Probability measures how likely an event is, on a scale from 0 to 1.

- For equally likely outcomes, $P(A) = \dfrac{\text{favourable outcomes}}{\text{total outcomes}}$.
- Events are **mutually exclusive** when they cannot both happen: $P(A \text{ or } B) = P(A) + P(B)$.
- Events are **independent** when one does not affect the other: $P(A \text{ and } B) = P(A) \times P(B)$.
- The **complement** rule, $P(\text{not } A) = 1 - P(A)$, is often the fastest route to an answer containing "at least one".
- **Counting techniques** supply the denominator: a **permutation** counts ordered selections, a **combination** counts unordered ones.

Conditional probability is $P(A \mid B) = P(A \cap B)/P(B)$ for $P(B) > 0$. A union subtracts its overlap: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Association in a supplied table does not by itself establish a cause.

Worked example. a second draw depends on the first



Known: two red and one blue counter, two random draws without replacement. To get two reds, follow the red-red path and multiply its conditional probabilities.

$$P(RR) = P(R_1)P(R_2 \mid R_1)$$

$$P(RR) = \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}$$

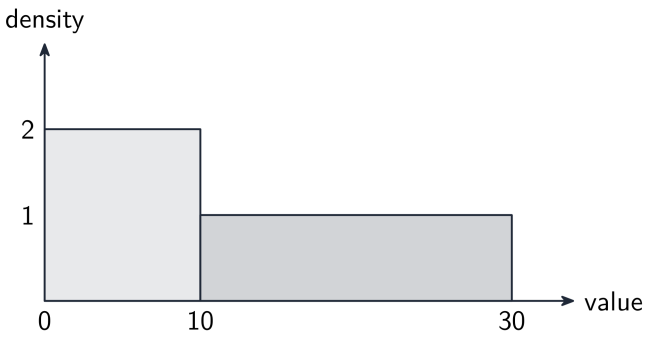
A first blue leaves only red counters; the blue-blue path has probability zero. The draws are not independent.

Practice sheet **2.4** gives a progressively harder set with worked solutions.

Statistics: collecting and displaying data

- A **population** is everyone you want to describe; a **sample** is who you actually measure.
- Generalising to a population needs a suitably **representative** sample and an honest account of selection and nonresponse. A class survey can describe that class without establishing a school-wide percentage.
- **Qualitative data** are categories; **quantitative data** measure amounts, either **discrete** (counted) or **continuous** (measured).
- **Bias** enters through who you ask, who answers, and how the question is worded, and arithmetic alone does not repair missing or systematically selected observations.

> **Worked example.** equal bar heights can conceal unequal frequencies



Known: class 0 to 10 has frequency 20; class 10 to 30 has frequency 20. Widths differ, so calculate frequency density.

$$D = \frac{f}{w}$$

$$D_1 = \frac{20}{10} = 2$$

$$D_2 = \frac{20}{20} = 1$$

The second bar is half as high but twice as wide. Both areas represent 20 observations.

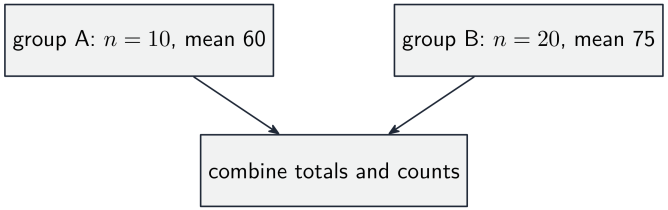
Practice sheet **2.5** gives a progressively harder set with worked solutions.

For unequal-width histogram classes, height is frequency divided by class width so that bar area represents frequency. Keep categories in separated bars and numeric continuous intervals in a histogram.

Statistics: measures of central location

- The **mean** is the total divided by the count. It uses every value, and every outlier.
- The **median** is the middle value in order. Moving an extreme value without changing the ordered middle may leave it unchanged.
- The **mode** is the most frequent value, and it can identify the most frequent category without imposing a numerical scale. A distribution may have more than one mode or no unique mode.
- △ Choose the summary for the question. A high extreme can pull the mean above the median. The mean still measures total per observation; the median locates the ordered middle. Neither says that most observations equal it.

> **Worked example.** group size determines a combined mean



Known: group A has 10 scores with mean 60; group B has 20 with mean 75. Recover each group's total before combining.

$$\bar{x} = \frac{n_A\bar{x}_A + n_B\bar{x}_B}{n_A + n_B}$$

$$\bar{x} = \frac{10(60) + 20(75)}{10 + 20} = 70$$

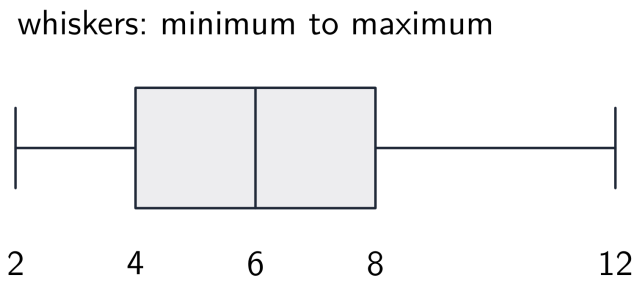
The unweighted mean of 60 and 75 would be 67.5, which ignores the unequal counts.

Practice sheet **2.6** gives a progressively harder set with worked solutions.

Statistics: measures of variability

- The **range** is the largest minus the smallest – the full span, so an extreme can change it substantially.
- The **interquartile range (IQR)** is the difference $Q_3 - Q_1$ spanning the central half by the stated quartile convention. A **box plot** shows quartiles and a median, with whisker rules that must be specified.
- The **standard deviation** summarises spread using squared deviations. For a complete population, use: $\sigma = \sqrt{\sum(x-\mu)^2/n}$. A sample estimate uses $s = \sqrt{\sum(x-\bar{x})^2/(n-1)}$ under that stated convention. Variance has squared units; standard deviation has the original units.
- A larger standard deviation means a wider spread. Two data sets can share a mean and describe completely different situations.

Worked example. read a five-number summary rather than every observation



Known five-number summary: minimum 2, $Q_1 = 4$, median 6, $Q_3 = 8$, maximum 12. The overall span and middle-half span are different.

$$R = x_{max} - x_{min} = 12 - 2 = 10$$

$$IQR = Q_3 - Q_1 = 8 - 4 = 4$$

The box does not identify every original value, the sample size or why the spread arose.

Practice sheet **2.7** gives a progressively harder set with worked solutions.

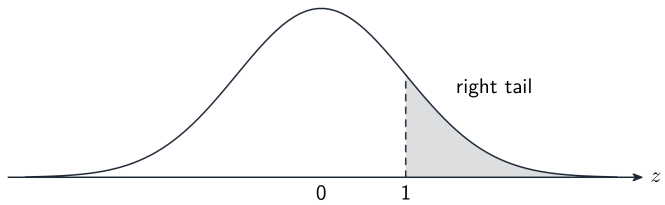
For sheet 2.7, quartiles are medians of ordered half-lists, excluding the overall median for an odd count. Whiskers run to the minimum and maximum. Do not mix that convention with software interpolated quartiles or outlier-fence whiskers.

Distributions: the normal distribution

- The **normal distribution** is a continuous symmetric bell-shaped model specified by its mean and positive standard deviation. Those two summaries alone do not establish that observed data follow this model.
- Under a normal model, the approximate empirical rule is: about 68% of values lie within one standard deviation of the mean, 95% within two, and 99.7% within three.
- A **z-score** says how many standard deviations a value sits from the mean: $z = \dfrac{x-\mu}{\sigma}$, which is what makes two different scales comparable.

For a normal model, use cumulative area $\Phi(z) = P(Z \le z)$ and symmetry $\Phi(-z) = 1 - \Phi(z)$. An interval subtracts its two cumulative bounds. Expected count is sample size times probability and need not be an integer; an observed count is an integer.

Worked example. a threshold is an area beyond a standardised value



Known: an invented measurement model has $\mu = 50$ and $\sigma = 10$. For values above 60, standardise the threshold.

$$z = \frac{x-\mu}{\sigma} = \frac{60-50}{10} = 1$$

Use the complement of the supplied cumulative area.

$$P(X > 60) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

This is a model probability, not a guarantee that exactly 15.87% of a finite sample exceeds 60.

Practice sheet **2.8** gives a progressively harder set with worked solutions.