# Unit S1 Statistics 1

GCE AS and GCE Mathematics, GCE AS and GCE Further Mathematics and GCE AS and GCE Further Mathematics (Additional) AS optional unit

## S1.1 Unit description

Mathematical models in probability and statistics; representation and summary of data; probability; correlation and regression; discrete random variables; discrete distributions; the Normal distribution.

## S1.2 Assessment information

**Examination**

The examination will consist of one 1½ hour paper. The paper will contain about seven questions with varying mark allocations per question which will be stated on the paper. All questions may be attempted.

**Calculators**

Students are expected to have available a calculator with at least the following keys: $+, -, \times, \div, \pi, x^2, \sqrt{x}, \frac{1}{x}, x^y, \ln x, e^x, x!$, sine, cosine and tangent and their inverses in degrees and decimals of a degree, and in radians; memory. Calculators with a facility for symbolic algebra, differentiation and/or integration are not permitted.

## Formulae

Students are expected to know formulae which might be required by the specification and which are not included in the booklet, Mathematical Formulae including Statistical Formulae and Tables, which will be provided for use with the paper. Questions will be set in SI units and other units in common usage.

Students will be expected to know and be able to recall and use the following formulae:

Mean = $\bar{x} = \frac{\Sigma x}{n}$ or $\frac{\Sigma fx}{\Sigma f}$

Standard deviation = $\sqrt{(\text{Variance})}$

Interquartile range = $\text{IQR} = \text{Q}_3 - \text{Q}_1$

$\text{P}(A') = 1 - \text{P}(A)$

For independent events $A$ and $B$,

$\text{P}(B \mid A) = \text{P}(B)$, $\text{P}(A \mid B) = \text{P}(A)$,

$\text{P}(A \cap B) = \text{P}(A)\ \text{P}(B)$

$\text{E}(aX + b) = a\text{E}(X) + b$

$\text{Var}\ (aX + b) = a^2\ \text{Var}\ (X)$

Cumulative distribution function for a discrete random variable:

$$\text{F}(x_0) = \text{P}(X \leq x0) = \sum_{x \leq x0} \text{p}(x)$$

Standardised Normal Random Variable $Z = \frac{X - \mu}{\sigma}$
where $X \sim \text{N}(\mu, \sigma^2)$

## 1 Mathematical models in probability and statistics

### What students need to learn:

The basic ideas of mathematical modelling as applied in probability and statistics.

## 2 Representation and summary of data

### What students need to learn:

| | |
|---|---|
| Histograms, stem and leaf diagrams, box plots. | Using histograms, stem and leaf diagrams and box plots to compare distributions.<br><br>Back-to-back stem and leaf diagrams may be required.<br><br>Drawing of histograms, stem and leaf diagrams or box plots will not be the direct focus of examination questions. |
| Measures of location — mean, median, mode. | Calculation of mean, mode and median, range and interquartile range will not be the direct focus of examination questions.<br><br>Students will be expected to draw simple inferences and give interpretations to measures of location and dispersion. Significance tests will not be expected.<br><br>Data may be discrete, continuous, grouped or ungrouped. Understanding and use of coding. |
| Measures of dispersion — variance, standard deviation, range and interpercentile ranges. | Simple interpolation may be required. Interpretation of measures of location and dispersion. |
| Skewness. Concepts of outliers. | Students may be asked to illustrate the location of outliers on a box plot. Any rule to identify outliers will be specified in the question. |

## 3 Probability

**What students need to learn:**

| | |
|---|---|
| Elementary probability. | |
| Sample space. Exclusive and complementary events. Conditional probability. | Understanding and use of<br>$P(A') = 1 - P(A)$,<br>$P(A \cup B) = P(A) + P(B) - P(A \cap B)$,<br>$P(A \cap B) = P(A)\,P(B \mid A)$. |
| Independence of two events. | $P(B \mid A) = P(B)$, $P(A \mid B) = P(A)$,<br>$P(A \cap B) = P(A)\,P(B)$. |
| Sum and product laws. | Use of tree diagrams and Venn diagrams. Sampling with and without replacement. |

## 4 Correlation and regression

**What students need to learn:**

| | |
|---|---|
| Scatter diagrams. Linear regression. | Calculation of the equation of a linear regression line using the method of least squares. Students may be required to draw this regression line on a scatter diagram. |
| Explanatory (independent) and response (dependent) variables. Applications and interpretations. | Use to make predictions within the range of values of the explanatory variable and the dangers of extrapolation. Derivations will not be required. Variables other than $x$ and $y$ may be used. Linear change of variable may be required. |
| The product moment correlation coefficient, its use, interpretation and limitations. | Derivations and tests of significance will not be required. |

  

## 5 Discrete random variables

**What students need to learn:**

| | |
|---|---|
| The concept of a discrete random variable. | |
| The probability function and the cumulative distribution function for a discrete random variable. | Simple uses of the probability function $p(x)$ where $p(x) = P(X = x)$. Use of the cumulative distribution function: $F(x_0) = P(X \leq x_0) = \sum_{x \leq x0} p(x)$ |
| Mean and variance of a discrete random variable. | Use of $E(X)$, $E(X^2)$ for calculating the variance of $X$. Knowledge and use of $E(aX + b) = aE(X) + b$, $Var(aX + b) = a^2 Var(X)$. |
| The discrete uniform distribution. | The mean and variance of this distribution. |

## 6 The Normal distribution

**What students need to learn:**

| | |
|---|---|
| The Normal distribution including the mean, variance and use of tables of the cumulative distribution function. | Knowledge of the shape and the symmetry of the distribution is required. Knowledge of the probability density function is not required. Derivation of the mean, variance and cumulative distribution function is not required. Interpolation is not necessary. Questions may involve the solution of simultaneous equations. |