5.12 AS MODULE 4732: PROBABILITY AND STATISTICS 1 (S1)

Preamble

Knowledge of the specification content of Modules *C1* and *C2* is assumed, and candidates may be required to demonstrate such knowledge in answering questions in Unit *S1*.

The specification content of this module includes some items that are part of GCSE Mathematics at Intermediate Tier (e.g. mean, mode, median, interquartile range, stem-and-leaf diagrams, box-and-whisker plots). These are included for completeness; where any such items are involved in examination questions, the main focus will be on interpretation rather than on elementary calculations.

The specification content of this module is to be understood in the context of modelling real-life situations, and examination questions may ask for comment and interpretation, including where appropriate, cross-checking between a model and reality.

Candidates should know the following formulae, none of which is included in the List of Formulae made available for use in the examination.

Representation of Data

Mean: $\dfrac{\Sigma x}{n}$ or $\dfrac{\Sigma xf}{\Sigma f}$

Standard deviation: $\sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\dfrac{\Sigma x^2}{n} - \bar{x}^2}$ or $\sqrt{\dfrac{\Sigma(x-\bar{x})^2 f}{\Sigma f}} = \sqrt{\dfrac{\Sigma x^2 f}{\Sigma f} - \bar{x}^2}$

Representation of Data

C3.1a C3.1b C3.2 C3.3 N3.1 N3.2 N3.3 IT3.3 WO3.1

Candidates should be able to:

(a) select a suitable way of presenting raw statistical data, and discuss advantages and/or disadvantages that particular representations of data may have (in addition to the representations in (c) below, candidates should be familiar with pie charts, bar charts and frequency polygons);

(b) extract from a table or statistical diagram salient features of the data, and express conclusions verbally;

(c) construct and interpret stem-and-leaf diagrams (including ordered and back-to-back stem-and-leaf diagrams), box-and-whisker plots, histograms and cumulative frequency graphs;

(d) understand, use and interpret different measures of central tendency (mean, median, mode) and variation (range, interquartile range, standard deviation), e.g. in comparing and contrasting sets of data;

(e) calculate the mean and standard deviation of a set of data (including grouped data) either from the data itself or from given totals such as Σx and Σx^2, or $\Sigma(x-a)$ and $\Sigma(x-a)^2$.

Probability

Candidates should be able to:

(a) understand the terms permutation and combination;

(b) solve problems about selections, e.g. finding the number of ways in which a team of 3 men and 2 women can be selected from a group of 6 men and 5 women;

(c) solve problems about arrangements of objects in a line, including those involving

  (i) repetition (e.g. the number of ways of arranging the letters of the word 'NEEDLESS'),

  (ii) restriction (e.g. the number of ways several people can stand in a line if 2 particular people must — or must not — stand next to each other);

(d) evaluate probabilities in simple cases by means of enumeration of elementary events (e.g. for the total score when two fair dice are thrown) or by calculation using permutations and combinations;

(e) use addition and multiplication of probabilities, as appropriate, in simple cases;

(f) understand informally the meaning of exclusive and independent events, and calculate and use conditional probabilities in simple cases, e.g. situations that can be represented by means of a tree diagram.

Discrete Random Variables

Candidates should be able to:

(a) construct a probability distribution table relating to a given situation involving a discrete random variable X, and calculate the expectation, variance and standard deviation of X (the notations $\mathrm{E}(X)$ for expectation (also referred to as expected value or mean) and $\mathrm{Var}(X)$ for variance, are included);

(b) use formulae for probabilities for the binomial and geometric distributions, and model given situations by one of these, as appropriate (the notations $\mathrm{B}(n, p)$ and $\mathrm{Geo}(p)$ are included);

(c) use tables of cumulative binomial probabilities (or equivalent calculator functions);

(d) use formulae for the expectation and variance of the binomial distribution, and for the expectation of the geometric distribution.

Bivariate Data

C3.1a C3.3 IT3.3

Candidates should be able to:

(a) calculate, both from simple raw data and from summarised data, the product moment correlation coefficient for a set of bivariate data;

(b) understand the basis of Spearman's coefficient of rank correlation, and calculate its value (questions set will not involve tied ranks);

(c) interpret the value of a product moment correlation coefficient or of Spearman's rank correlation coefficient in relation to the appearance of a scatter diagram, with particular reference to values close to $-1, 0, 1$;

(d) understand that the value of a correlation coefficient is unaffected by linear transformations (coding) of the variables;

(e) understand the difference between an independent (or controlled) variable and a dependent variable;

(f) understand the concepts of least squares and regression lines in the context of a scatter diagram;

(g) calculate, both from simple raw data and from summarised data, the equation of a regression line, understand the distinction between the regression line of y on x and that of x on y, and use the fact that both regression lines pass through the mean centre $(\bar{x}, \bar{y})$;

(h) select and use, in the context of a problem, the appropriate regression line to estimate a value, and be able to interpret in context the uncertainties of such estimations.