5.14 A2 MODULE 4734: PROBABILITY AND STATISTICS 3 (S3)

Preamble

Knowledge of the specification content of Modules *C1* to *C4*, *S1* and *S2* is assumed, and candidates may be required to demonstrate such knowledge in answering questions in Unit *S3*.

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

Expectation Algebra

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

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

$\mathrm{E}(aX + bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y)$

Continuous Random Variables

Candidates should be able to:

(a) use probability density functions which may be defined ‘piecewise’;

(b) use, in simple cases, the general result $\mathrm{E}(\mathrm{g}(X)) = \int \mathrm{g}(x)\,\mathrm{f}(x)\,\mathrm{d}x$, where $\mathrm{f}(x)$ is the probability density function of the continuous random variable *X* and $\mathrm{g}(X)$ is a function of *X*;

(c) understand and use the relationship between the probability density function and the (cumulative) distribution function, and use either to evaluate the median, quartiles and other percentiles;

(d) use (cumulative) distribution functions of related variables in simple cases, e.g. given the c.d.f. of a variable *X*, to find the c.d.f. and hence the p.d.f. of *Y*, where $Y = X^3$.

Linear Combinations of Random Variables

Candidates should be able to:

(a) use, in the course of solving problems, the results that

(i) $\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b$ and $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$,

(ii) $\mathrm{E}(aX + bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y)$,

(iii) $\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)$ for independent *X* and *Y*,

(iv) if *X* has a normal distribution then so does $aX + b$,

(v) if *X* and *Y* have independent normal distributions then $aX + bY$ has a normal distribution,

(vi) if *X* and *Y* have independent Poisson distributions then $X + Y$ has a Poisson distribution.

Confidence Intervals; the *t* Distribution

Candidates should be able to:

(a) determine a confidence interval for a population mean, using a normal distribution, in the context of

- (i) a sample drawn from a normal population of known variance,
- (ii) a large sample, using the Central Limit Theorem and an unbiased variance estimate derived from the sample;

(b) determine, from a large sample, an approximate confidence interval for a population proportion;

(c) use a *t* distribution, with the appropriate number of degrees of freedom, in the context of a small sample drawn from a normal population of unknown variance

- (i) to determine a confidence interval for the population mean,
- (ii) to carry out a hypothesis test of the population mean.

Difference of Population Means and Proportions

Candidates should be able to:

(a) understand the difference between a two-sample test and a paired-sample test, and select the appropriate form of test in solving problems;

(b) formulate hypotheses and carry out a test for a difference of population means or population proportions using a normal distribution, and appreciate the conditions necessary for the test to be valid;

(c) calculate a pooled estimate of a population variance based on the data from two samples;

(d) formulate hypotheses and carry out either a two-sample *t*-test or a paired-sample *t*-test, as appropriate, for a difference of population means, and appreciate the conditions necessary for these tests to be valid;

(e) calculate a confidence interval for a difference of population means, using a normal distribution or a *t* distribution, as appropriate.

χ^2 tests

Candidates should be able to:

(a) fit a theoretical distribution, as prescribed by a given hypothesis, to given data (questions set will not involve lengthy calculations);

(b) use a χ^2 test with the appropriate number of degrees of freedom to carry out the corresponding goodness of fit test (classes should be combined so that each expected frequency is at least 5);

(c) use a χ^2 test with the appropriate number of degrees of freedom to test for independence in a contingency table (rows or columns, as appropriate, should be combined so that each expected frequency is at least 5, and Yates' correction should be used in the special case of a 2×2 table).