



**ADVANCED GCE UNIT
MATHEMATICS (MEI)**

4769/01

Statistics 4

TUESDAY 5 JUNE 2007

Afternoon

Time: 1 hour 30 minutes

Additional Materials:

Answer booklet (8 pages)

Graph paper

MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer any **three** questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [] at the end of each question or part question.
- The total number of marks for this paper is 72.

ADVICE TO CANDIDATES

- Read each question carefully and make sure you know what you have to do before starting your answer.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

This document consists of **4** printed pages.

Option 1: Estimation

1 The random variable X has the continuous uniform distribution with probability density function

$$f(x) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta,$$

where θ ($\theta > 0$) is an unknown parameter.

A random sample of n observations from X is denoted by X_1, X_2, \dots, X_n , with sample mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

(i) Show that $2\bar{X}$ is an unbiased estimator of θ . [4]

(ii) Evaluate $2\bar{X}$ for a case where, with $n = 5$, the observed values of the random sample are 0.4, 0.2, 1.0, 0.1, 0.6. Hence comment on a disadvantage of $2\bar{X}$ as an estimator of θ . [4]

For a general random sample of size n , let Y represent the sample maximum, $Y = \max(X_1, X_2, \dots, X_n)$. You are given that the probability density function of Y is

$$g(y) = \frac{ny^{n-1}}{\theta^n}, \quad 0 \leq y \leq \theta.$$

(iii) An estimator kY is to be used to estimate θ , where k is a constant to be chosen. Show that the mean square error of kY is

$$k^2 E(Y^2) - 2k\theta E(Y) + \theta^2$$

and hence find the value of k for which the mean square error is minimised. [12]

(iv) Comment on whether kY with the value of k found in part **(iii)** suffers from the disadvantage identified in part **(ii)**. [4]

Option 2: Generating Functions

2 The random variable X has the binomial distribution with parameters n and p , i.e. $X \sim B(n, p)$.

(i) Show that the probability generating function of X is $G(t) = (q + pt)^n$, where $q = 1 - p$. [4]

(ii) Hence obtain the mean μ and variance σ^2 of X . [6]

(iii) Write down the mean and variance of the random variable $Z = \frac{X - \mu}{\sigma}$. [1]

(iv) Write down the moment generating function of X and use the linear transformation result to show that the moment generating function of Z is

$$M_Z(\theta) = \left(qe^{-\frac{p\theta}{\sqrt{npq}}} + pe^{\frac{q\theta}{\sqrt{npq}}} \right)^n. \quad [5]$$

(v) By expanding the exponential terms in $M_Z(\theta)$, show that the limit of $M_Z(\theta)$ as $n \rightarrow \infty$ is $e^{\theta^2/2}$.

You may use the result $\lim_{n \rightarrow \infty} \left(1 + \frac{y + f(n)}{n} \right)^n = e^y$ provided $f(n) \rightarrow 0$ as $n \rightarrow \infty$. [4]

(vi) What does the result in part (v) imply about the distribution of Z as $n \rightarrow \infty$? Explain your reasoning briefly. [3]

(vii) What does the result in part (vi) imply about the distribution of X as $n \rightarrow \infty$? [1]

Option 3: Inference

3 An engineering company buys a certain type of component from two suppliers, A and B. It is important that, on the whole, the strengths of these components are the same from both suppliers. The company can measure the strengths in its laboratory. Random samples of seven components from supplier A and five from supplier B give the following strengths, in a convenient unit.

Supplier A 25.8 27.4 26.2 23.5 28.3 26.4 27.2

Supplier B 25.6 24.9 23.7 25.8 26.9

The underlying distributions of strengths are assumed to be Normal for both suppliers, with variances 2.45 for supplier A and 1.40 for supplier B.

(i) Test at the 5% level of significance whether it is reasonable to assume that the mean strengths from the two suppliers are equal. [10]

(ii) Provide a two-sided 90% confidence interval for the true mean difference. [4]

(iii) Show that the test procedure used in part (i), with samples of sizes 7 and 5 and a 5% significance level, leads to acceptance of the null hypothesis of equal means if $-1.556 < \bar{x} - \bar{y} < 1.556$, where \bar{x} and \bar{y} are the observed sample means from suppliers A and B. Hence find the probability of a Type II error for this test procedure if in fact the true mean strength from supplier A is 2.0 units more than that from supplier B. [7]

(iv) A manager suggests that the Wilcoxon rank sum test should be used instead, comparing the median strengths for the samples of sizes 7 and 5. Give one reason why this suggestion might be sensible and two why it might not. [3]

Option 4: Design and Analysis of Experiments

- 4 An agricultural company conducts a trial of five fertilisers (A, B, C, D, E) in an experimental field at its research station. The fertilisers are applied to plots of the field according to a completely randomised design. The yields of the crop from the plots, measured in a standard unit, are analysed by the one-way analysis of variance, from which it appears that there are no real differences among the effects of the fertilisers.

A statistician notes that the residual mean square in the analysis of variance is considerably larger than had been anticipated from knowledge of the general behaviour of the crop, and therefore suspects that there is some inadequacy in the design of the trial.

- (i) Explain briefly why the statistician should be suspicious of the design. [2]
- (ii) Explain briefly why an inflated residual leads to difficulty in interpreting the results of the analysis of variance, in particular that the null hypothesis is more likely to be accepted erroneously. [3]

Further investigation indicates that the soil at the west side of the experimental field is naturally more fertile than that at the east side, with a consistent 'fertility gradient' from west to east.

- (iii) What experimental design can accommodate this feature? Provide a simple diagram of the experimental field indicating a suitable layout. [4]

The company decides to conduct a new trial in its glasshouse, where experimental conditions can be controlled so that a completely randomised design is appropriate. The yields are as follows.

Fertiliser A	Fertiliser B	Fertiliser C	Fertiliser D	Fertiliser E
23.6	26.0	18.8	29.0	17.7
18.2	35.3	16.7	37.2	16.5
32.4	30.5	23.0	32.6	12.8
20.8	31.4	28.3	31.4	20.4

[The sum of these data items is 502.6 and the sum of their squares is 13 610.22.]

- (iv) Construct the usual one-way analysis of variance table. Carry out the appropriate test, using a 5% significance level. Report briefly on your conclusions. [12]
- (v) State the assumptions about the distribution of the experimental error that underlie your analysis in part (iv). [3]