General Certificate of Education
June 2004
Advanced Level Examination

MATHEMATICS AND STATISTICS   **MBS4**
(SPECIFICATION B)
Unit Statistics 4

Wednesday 16 June 2004   Afternoon Session

In addition to this paper you will require:
  • a 12-page answer book;
  • the AQA booklet of formulae and statistical tables.
You may use a graphics calculator.

Time allowed: 1 hour 45 minutes

Instructions
• Use blue or black ink or ball-point pen. Pencil should only be used for drawing.
• Write the information required on the front of your answer book. The *Examining Body* for this
  paper is AQA. The *Paper Reference* is MBS4.
• Answer **all** questions.
• All necessary working should be shown; otherwise marks for method may be lost.
• The **final** answer to questions requiring the use of tables or calculators should normally be given to
  three significant figures.

Information
• The maximum mark for this paper is 80.
• Mark allocations are shown in brackets.

Advice
• Unless stated otherwise, formulae may be quoted, without proof, from the booklet.

MBS4

Answer **all** questions.

1 Majid delivers milk. His milk round consists of 280 houses which he visits each day. The daily number of pints of milk requested at each house may be modelled by the random variable, X, with the following probability distribution.

x	0	1	2	3	4	5
P($X = x$)	0.08	0.30	0.34	0.15	0.10	0.03

  (a) Calculate the mean and standard deviation of X. *(6 marks)*

(b)   (i) Find the mean and standard deviation of $\overline{X}$, the mean daily number of pints of milk per house requested at the 280 houses on his round. *(3 marks)*

  (ii) Find k, such that the probability that $\overline{X}$ is greater than k is approximately 0.01. *(3 marks)*

  (iii) How many pints of milk should Majid take with him on his round in order that the probability that he runs out of milk before completing his round is approximately 0.01? *(1 mark)*

2 Priscilla supplies chilled food cabinets for use when displaying perishable foods in shops. She claims that the mean time to spoilage of yoghurt displayed in one of her cabinets is more than 200 hours.

To test this, Rio, a shopkeeper, placed 11 randomly selected tubs of fresh yoghurt in one of Priscilla's cabinets with a 'Not for Sale' sign. The times to spoilage, in hours, of the yoghurt in these tubs were as follows.

206   213   197   215   202   200   209   217   195   203   204

(a) Test Priscilla's claim using the 5% significance level. You may assume that the data is a sample from a normal distribution. *(10 marks)*

(b) Discuss the problems, if any, which would have arisen in the interpretation of the results of the test if three of the tubs of yoghurt had accidentally been sold and so data were only available for the remaining eight tubs. *(2 marks)*

(c) Define a Type I error in the context of this question. *(2 marks)*

(d) Adam, another shopkeeper, intends to repeat Rio's data collection and analysis. State the probability of a Type I error occurring in the analysis of Adam's data when the mean time to spoilage of the tubs of fresh yoghurt displayed in Priscilla's cabinets is:

  (i) 200 hours;

  (ii) 202 hours. *(3 marks)*

TURN OVER FOR THE NEXT QUESTION

3 A company manufactures components for the motor industry. The components are designed to have a length of 135.0 mm. Each day a technician takes a random sample of components and calculates a 95% confidence interval for the mean length.

(a) On a particular Monday the technician takes a sample of nine components. Their lengths, in millimetres, are as follows.

135.1   135.7   134.9   135.2   136.3   135.7   135.9   136.0   135.6

Assuming the lengths of components produced on this Monday may be modelled by a normal distribution, calculate a 95% confidence interval for the mean length. Give your answer to an appropriate degree of accuracy. *(8 marks)*

(b) On a particular Friday a random sample of 60 components had a mean length of 135.8 mm and a standard deviation of 3.9 mm. Calculate a 95% confidence interval for the mean length of components produced on this Friday. *(4 marks)*

(c) The company overhauls all machines when the confidence interval calculated does not contain 135.0 mm. State whether increasing the sample size, *n*, is likely to increase, decrease or leave unaltered:

(i) the width of the confidence interval; *(1 mark)*

(ii) the frequency of machine overhauls, explaining your answer. *(2 marks)*

(d) Using your results in parts (a) and (b) as an example, or otherwise, explain why it is possible that production on a day when the confidence interval does not contain 135.0 mm will sometimes be more satisfactory than production on a day when the confidence interval does contain 135.0 mm. *(2 marks)*

4 The random variable, X, has probability density function

$$f(x) = \begin{cases} cx + d & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases} \qquad \text{where } c \text{ and } d \text{ are constants}$$

(a) Show that $2c + 2d = 1$. *(3 marks)*

(b) Find the median of X:

   (i) when $c = 0$ and $d = 0.5$; *(2 marks)*

   (ii) when $c = 0.5$ and $d = 0$. *(3 marks)*

(c) The following pairs of values are suggested for c and d.

   Suggestion A: $c = 0.5$ and $d = 0.5$
   Suggestion B: $c = -0.5$ and $d = 1$
   Suggestion C: $c = 1.5$ and $d = -1$

   Only one of the suggestions provides a valid probability density function.

   For **each** suggestion state whether or not it is valid. If you state it is not valid, give a reason. *(5 marks)*

TURN OVER FOR THE NEXT QUESTION

5 The board of an insurance company was concerned about the poor performance of its motor insurance business. Belinda was asked to collect and analyse data on the claims record of the 4902 customers who held motor policies in 2003.

(a) The following contingency table summarises the number of claims made in 2003 classified by length of time the policy had been held.

		Number of Claims in 2003			
		0	**1**	**2**	**3 or more**
Policy held for	**less than 2 years**	1478	212	21	6
	2 or more years	2830	334	20	1

Use a χ^2 distribution and the 1% significance level to investigate whether the number of claims made in 2003 is associated with the length of time for which the policy had been held. *(11 marks)*

(b) Belinda also constructed the following tables concerning the 4902 customers.

Table A

		% of Total Number of Policy Holders who in 2003	
		Made a Claim	**Did not make a Claim**
Gender	**Male**	8.4	57.6
	Female	3.7	30.3

Table B

		Total Number of Policy Holders in 2003	**Number of Policy Holders who made a claim in 2003**
Age (years)	**<25**	956	280
	25 or more	3946	314

Table C

		Number of Policy Holders who in 2003	
		Did not make a Claim	**Made a claim for more than £2000**
Age (years)	**<25**	676	145
	25 or more	3632	192

For **each** table:

(i) state why, in its present form, it cannot be analysed as a contingency table using a χ^2 distribution; *(3 marks)*

(ii) construct a new table which can be analysed as a contingency table using a χ^2 distribution. **Do not analyse any of the tables you construct**. *(6 marks)*

END OF QUESTIONS

THERE ARE NO QUESTIONS PRINTED ON THIS PAGE