General Certificate of Education
June 2009
Advanced Level Examination

MATHEMATICS **MS04**
Unit Statistics 4

Tuesday 23 June 2009 9.00 am to 10.30 am

For this paper you must have:
- an 8-page answer book
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Time allowed: 1 hour 30 minutes

Instructions
- Use black ink or black 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 MS04.
- Answer **all** questions.
- Show all necessary working; 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 75.
- The marks for questions are shown in brackets.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.

Answer **all** questions.

1 A teacher believes that the Verbal Reasoning Quotient (VRQ) of first-born children in a family is higher than that of subsequent children in the family. The teacher randomly selects ten families, which each have two children, and then records the VRQ for each child. The results are shown in the table.

Family	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
First-born VRQ	110	108	99	107	121	128	107	110	123	106
Second-born VRQ	110	108	100	101	123	127	103	106	122	103

Assuming that differences in the VRQ are normally distributed, investigate the teacher's belief at the 5% level of significance. *(10 marks)*

2 The random variable X denotes the number of trials necessary in order to obtain the first success.

(a) State **three** conditions which must apply in order that X may be modelled by a geometric distribution. *(3 marks)*

(b) The discrete random variable Y is such that $Y \sim \text{Geo}(p)$ and $\text{P}(Y = 1 \text{ or } 2) = 0.2775$.

(i) Determine the value of p. *(4 marks)*

(ii) Hence find the values of $\text{E}(Y)$ and $\text{Var}(Y)$. *(2 marks)*

3 Fourteen randomly selected physics students are asked to determine independently the density of copper by a particular method. Their results, in kg m^{-3}, are shown below.

8924 8929 8931 8926 8928 8925 8929

8921 8925 8927 8930 8920 8923 8922

(a) Construct a 98% confidence interval for the standard deviation of the density of copper as determined by a physics student using this particular method. *(6 marks)*

(b) State **one** assumption that you have made in constructing this confidence interval. *(1 mark)*

4 The rateable values of businesses in a certain town are distributed with mean μ and variance σ^2. Two trainee valuers are asked to estimate μ.

Trainee valuer A intends to take a random sample of 15 businesses and to calculate the mean rateable value, $\overline{X}_A$, of the businesses in this sample.

(a) Write down expressions for the mean and the variance of $\overline{X}_A$. *(2 marks)*

(b) Trainee valuer B intends to take a random sample of 10 businesses and to calculate the mean rateable value, $\overline{X}_B$, of the businesses in this sample.

The principal valuer suggests that they could obtain a better estimate of μ by combining their individual estimators.

The trainee valuers consider two possible combinations, $\overline{X}_L$ and $\overline{X}_M$, of their estimators, where

$$\overline{X}_L = \tfrac{1}{2}\overline{X}_A + \tfrac{1}{2}\overline{X}_B \quad \text{and} \quad \overline{X}_M = \tfrac{3}{5}\overline{X}_A + \tfrac{2}{5}\overline{X}_B$$

(i) Show that the estimator $\overline{X}_M$ has mean μ and variance $\dfrac{\sigma^2}{25}$. *(3 marks)*

(ii) Show that $\overline{X}_L$ is an unbiased estimator of μ. *(1 mark)*

(iii) Calculate the relative efficiency of $\overline{X}_L$ with respect to $\overline{X}_M$, and give a reason why $\overline{X}_M$ should be preferred to $\overline{X}_L$ as an estimator of μ. *(5 marks)*

5 Henrietta, a statistician, buys boxes of 6 eggs from a local farm shop. She believes that the number of light brown eggs in each box follows a binomial distribution. In order to test this belief, she selects a random sample of 100 boxes of eggs from the farm shop and records the number of light brown eggs in each box, with the following results.

Number of light brown eggs per box	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Frequency	23	32	23	17	4	1	0	100

(a) (i) Use these data to estimate the mean number of light brown eggs in a box of 6 eggs. *(1 mark)*

(ii) Hence show that a suitable estimate of p, the probability that a randomly chosen egg is light brown, is 0.25. *(1 mark)*

(b) By calculating an appropriate χ^2-statistic, test, at the 5% level of significance, whether a binomial distribution is an appropriate model for the number of light brown eggs in a box of 6 eggs. *(10 marks)*

Turn over for the next question

6 The heights in centimetres, X, of men from police forces in England may be assumed to be normally distributed with variance ${\sigma_X}^2$. The heights in centimetres, Y, of men from police forces in Scotland may also be assumed to be normally distributed but with variance ${\sigma_Y}^2$.

The heights, measured to the nearest centimetre, x, of a random sample of 11 men from police forces in England were

179 189 191 179 188 178 189 181 174 186 187

The heights, measured to the nearest centimetre, y, of a random sample of 9 men from police forces in Scotland were

182 180 178 175 177 180 182 181 181

(a) Calculate unbiased estimates of ${\sigma_X}^2$ and ${\sigma_Y}^2$. *(2 marks)*

(b) (i) Hence determine a 95% confidence interval for the variance ratio $\dfrac{{\sigma_X}^2}{{\sigma_Y}^2}$. *(7 marks)*

(ii) Comment on the suggestion that the heights of men from police forces in England are more variable than those of men from police forces in Scotland. *(2 marks)*

7 The continuous random variable X is modelled by an exponential distribution with probability density function

$$\mathrm{f}(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

(a) Write down the cumulative distribution function, $\mathrm{F}(x)$. *(2 marks)*

(b) Show that the **exact** value of the interquartile range of X is given by $\dfrac{1}{\lambda}\ln 3$. *(5 marks)*

(c) (i) Use integration to prove that $\mathrm{E}(X^2) = \dfrac{2}{\lambda^2}$. *(4 marks)*

(ii) Hence, given that $\mathrm{E}(X) = \dfrac{1}{\lambda}$, show that $\mathrm{Var}(X) = \dfrac{1}{\lambda^2}$. *(1 mark)*

(d) (i) Find the **exact** value of λ for which the value of the interquartile range of X is four times the value of the variance of X. *(2 marks)*

(ii) Describe what happens to the interquartile range of X as λ increases without bound. *(1 mark)*

END OF QUESTIONS