AQA

Please write clearly in block capitals.

Centre number | | | | | | Candidate number | | | | |

Surname ______________________________

Forename(s) ______________________________

Candidate signature ______________________________

A-level
MATHEMATICS
Unit Statistics 4

Tuesday 26 June 2018 Morning Time allowed: 1 hour 30 minutes

Materials
For this paper you must have:
- the blue AQA booklet of formulae and statistical tables.

You may use a graphics calculator.

Instructions
- Use black ink or black ball-point pen. Pencil should only be used for drawing.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- Write the question part reference (eg (a), (b)(i) etc) in the left-hand margin.
- You must answer each question in the space provided for that question. If you require extra space, use an AQA supplementary answer book; do **not** use the space provided for a different question.
- Do not write outside the box around each page.
- Show all necessary working; otherwise marks for method may be lost.
- Do all rough work in this book. Cross through any work that you do not want to be marked.
- The **final** answer to questions requiring the use of tables or calculators should normally be given to three significant figures.

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
TOTAL	

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 75.

Advice
- Unless stated otherwise, you may quote formulae, without proof, from the booklet.
- You do not necessarily need to use all the space provided.

Answer **all** questions.

Answer each question in the space provided for that question.

1 The table shows, for each of a random sample of six sets of non-identical twin girls, the heights, in centimetres, at age 25 years, of the first born and of the second born.

		Set of twin girls					
		A	**B**	**C**	**D**	**E**	**F**
Height at age 25 years	**First born**	156.7	150.8	162.4	166.2	163.6	170.2
	Second born	153.0	153.6	168.2	169.2	167.1	166.0

Stating a necessary assumption, construct a 90% confidence interval for the difference in mean height, at age 25 years, between first and second born non-identical twin girls.

[6 marks]

QUESTION PART REFERENCE

Answer space for question 1

QUESTION PART REFERENCE	Answer space for question 1

Turn over ▶

2 For many years, a company has been producing plastic pipes of nominal length one metre. During this time, it has been established that the lengths of the pipes produced have a mean of $1015\,\mathrm{mm}$ and a standard deviation of $10\,\mathrm{mm}$.

Following the installation of a new process for producing the pipes, a random sample of 16 pipes had the following lengths, in millimetres.

1011	1004	1012	1003	997	1010	1016	1000
1014	1018	996	1011	1015	1017	1006	1014

Assume that the lengths of plastic pipes produced by the new process can be modelled by the distribution $\mathrm{N}(\mu,\, \sigma^2)$.

(a) Test, at the 5% level of significance, the claim that the standard deviation of the lengths of pipes produced by the new process is no different to that established previously.

[8 marks]

(b) Based on your conclusion in part **(a)**, indicate the type of hypothesis test that you would carry out in order to test $\mathrm{H}_0 : \mu = 1015$ against $\mathrm{H}_1 : \mu < 1015$. Justify your answer but do **not** carry out the test.

[2 marks]

QUESTION PART REFERENCE

Answer space for question 2

Do not write outside the box

QUESTION PART REFERENCE	Answer space for question 2

Turn over ▶

Do not write outside the box

3 The random variable X has an exponential distribution with mean $\mu = \frac{1}{\lambda}$.

(a) (i) Use integration to find an expression, in terms of λ, for the median, m, of X.

[4 marks]

(ii) Evaluate $P(m < X < \mu)$.

[3 marks]

(b) The interval, Y days, between successive calls from a village telephone box can be modelled by the following cumulative distribution function.

$$F(y) = \begin{cases} 0 & y < 0 \\ 1 - e^{-0.0125y} & 0 \leqslant y < \infty \end{cases}$$

Calculate:

(i) $P(Y < 2E(Y))$;

[2 marks]

(ii) the probability that no calls will be made from the telephone box during 2019. Assume that the above model for Y is valid throughout 2019.

[3 marks]

QUESTION PART REFERENCE	Answer space for question 3

Do not write outside the box

QUESTION PART REFERENCE

Answer space for question 3

Turn over ▶

QUESTION PART REFERENCE	Answer space for question 3

Do not write outside the box

QUESTION PART REFERENCE	**Answer space for question 3**

Turn over ▶

Do not write outside the box

4 In the manufacture of 100-metre balls of string, it is claimed that faults occur at random and at a constant mean rate of λ per ball. The value of λ is unknown.

A random sample of 125 balls, each containing 100 metres of string, is monitored for faults during the winding process. The results are shown in the table.

Number of faults	0	1	2	3	4	5	6	$\geqslant 7$
Number of balls	14	23	35	23	16	10	4	0

(a) Show that an estimate of λ is 2.4 .

[1 mark]

(b) Hence test, at the 5% level of significance, the claim that faults occur at random and at a constant mean rate.

[10 marks]

QUESTION PART REFERENCE

Answer space for question 4

Do not write outside the box

QUESTION PART REFERENCE

Answer space for question 4

Turn over ▶

QUESTION PART REFERENCE	Answer space for question 4

QUESTION PART REFERENCE	Answer space for question 4

Do not write outside the box

5 Every evening after work, Carla takes her dog for a walk. Carla chooses either a rural walk or an urban walk.

Over a period of time, Carla recorded, in minutes, the durations of her walks.

A summary of her recorded durations, which may be assumed to constitute independent random samples, is shown in the table.

		Sample		
		Size	**Mean**	**Variance** (s^2)
Walk	**Rural**	11	36.0	51.40
	Urban	21	32.0	32.40

(a) Use a hypothesis test to confirm that there is no evidence, at the 5% level of significance, of a difference between the variability in duration of Carla's evening rural walks and that of her evening urban walks.

[6 marks]

(b) (i) Hence construct a 95% confidence interval for the difference between the mean duration of Carla's evening rural walks and that of her evening urban walks. Give the limits to one decimal place.

[6 marks]

(ii) State, with justification, what can be concluded from your answer to part **(b)(i)**.

[2 marks]

QUESTION PART REFERENCE

Answer space for question 5

QUESTION PART REFERENCE

Answer space for question 5

Do not write outside the box

QUESTION PART REFERENCE	Answer space for question 5

Do not write outside the box

QUESTION PART REFERENCE

Answer space for question 5

Turn over ▶

6 (a) The random variable X has a geometric distribution with parameter p.

(i) Prove that $\mathrm{E}(X) = \dfrac{1}{p}$.

[3 marks]

(ii) Given that $\mathrm{E}(X(X-1)) = \dfrac{2(1-p)}{p^2}$, deduce an expression for $\mathrm{Var}(X)$.

[2 marks]

(iii) Prove that $\displaystyle\sum_{x=n}^{\infty} \mathrm{P}(X = x) = (1-p)^{n-1}$, where n is an integer.

[2 marks]

(b) A spinning wheel consists of 37 equally-sized slots numbered $0, 1, 2, 3, \ldots, 36$. On any spin of the wheel, the score obtained is equal to the number of the slot into which a ball falls.

The random variable Y denotes the number of spins of such a wheel necessary to first obtain a score of 0.

(i) Find values for $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$.

[2 marks]

(ii) Find the minimum value of n such that $\mathrm{P}(Y \geqslant n) < 0.01$.

[3 marks]

QUESTION PART REFERENCE	**Answer space for question 6**

QUESTION PART REFERENCE	Answer space for question 6

Do not write outside the box

QUESTION PART REFERENCE	**Answer space for question 6**

QUESTION PART REFERENCE

Answer space for question 6

7 The independent random variables X and Y have Poisson distributions with parameters λ and 4λ respectively.

The variable S denotes the sum of a random sample of $\boldsymbol{n}$ observations on X.
The variable T denotes the sum of a random sample of $\mathbf{2}\boldsymbol{n}$ observations on Y.

The following two estimators are proposed for λ.

$$U = \frac{1}{9n}(S+T) \qquad V = \frac{1}{12n}(4S+T)$$

(a) Show that U and V are unbiased estimators for λ.

[3 marks]

(b) (i) Derive simplified expressions, in terms of λ and n, for each of $\mathrm{Var}(U)$ and $\mathrm{Var}(V)$.

[4 marks]

(ii) Hence state why U and V are consistent estimators for λ.

[1 mark]

(c) Calculate the efficiency of U relative to V.

[2 marks]

QUESTION PART REFERENCE

Answer space for question 7

Do not write outside the box

QUESTION PART REFERENCE	Answer space for question 7

Turn over ▶

Do not write outside the box

QUESTION PART REFERENCE	Answer space for question 7

Do not write outside the box

QUESTION PART REFERENCE	Answer space for question 7

END OF QUESTIONS

There are no questions printed on this page

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

Do not write outside the box

There are no questions printed on this page

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

Do not write outside the box

There are no questions printed on this page

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED