Please write clearly in block capitals.

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

Surname _____

Forename(s) _____

Candidate signature _____

# A-level
# STATISTICS

## Unit Statistics 4

Tuesday 19 June 2018 Afternoon 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 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.

**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.

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

Do not write outside the box

Answer **all** questions.

Answer each question in the space provided for that question.

**1** Li Wei is the chief librarian of a public library. As part of a study into library use, he is interested in the times that users spend in the library.

For his study, Li Wei measures the time, $x$ minutes, spent in the library by each of a random sample of $25$ users. He finds that:

$$\sum x = 478.5 \quad \text{and} \quad \sum (x - \bar{x})^2 = 698.91$$

Assume that times spent by users in the library may be modelled by a normal distribution.

**(a)** Construct a $99\%$ confidence interval for the mean time that users spend in the library.

**[5 marks]**

**(b)** Li Wei thinks that the interval is too wide to include in a report that he is writing so he decides to take another sample. He still wants to construct a $99\%$ confidence interval.

His assistant, Ranesh, has assured him that if he calculates his interval from a larger sample, then he will certainly end up with a narrower interval than the one constructed in part **(a)**.

Explain why this is not necessarily the case.

**[2 marks]**

| QUESTION PART REFERENCE | **Answer space for Question 1** |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 1 |
|---|---|
| | |

**Turn over ▶**

Do not write outside the box

**2** The probability that an egg laid by a hen will have a double yolk depends on the age of the hen. For a young hen that has just started laying eggs, this probability is $0.01$ while for an older hen at the end of her laying life, the probability is $0.001$.

**(a)** One egg from a young hen and one egg from an older hen are selected at random.

Find the probability that exactly one of these two eggs has a double yolk.

**[2 marks]**

**(b)** A random sample of $40$ eggs laid by young hens is taken.

Use a binomial distribution to find the exact probability that more than one of these eggs will have a double yolk.

**[2 marks]**

**(c)** A random sample of $400$ eggs laid by older hens is taken.

Use a distributional approximation to estimate the probability that at least one but fewer than four of these eggs will have double yolks. Give your answer to **four** decimal places.

You should state the distributional approximation that you use.

**[4 marks]**

**(d)** An egg distributor is supplied with eggs from two farms, A and B. Farm A has mainly young hens while farm B has mainly older hens.

The distributor packs the eggs onto trays of $30$ eggs by selecting $15$ eggs from each farm.

State whether a binomial distribution is a reasonable model for the number of eggs with a double yolk on a tray. Justify your answer in the context of the question.

**[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

| 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** Elaine is the manager of a single-screen cinema. The second film in a proposed series of *Martian Zombies* films has just opened at her cinema.

It has been reported that, in general, $20$ per cent of people who go to watch the second film in a series have **not** seen the first film in the series.

Elaine wishes to investigate whether this percentage applies to her customers for *Martian Zombies 2*. So, at a particular screening of this film, she asks a random sample of $20$ customers whether they have seen the first *Martian Zombies* film.

**(a) (i)** She finds that $8$ of the $20$ customers she asked have not seen the first film.

Use an exact test at the $5\%$ level of significance to decide whether the reported value of $20$ per cent applies to Elaine's customers for *Martian Zombies 2*.

**[6 marks]**

**(ii)** If Elaine had actually found that $9$ of the $20$ customers she asked had not seen the first film, find the smallest level of significance at which the null hypothesis for the test carried out in part **(a)(i)** would have been rejected.

**[3 marks]**

**(b)** Elaine's assistant, Roger, thinks that her sample of $20$ customers is too small.

So, at the cinema's next screening of *Martian Zombies 2*, Roger asks all $105$ customers whether they have seen the first *Martian Zombies* film.

Do you think the data collected by Roger is suitable for applying the test you have used in part **(a)**? Explain your answer in context.

**[2 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 ▶**

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 3 |
|---|---|
| | |

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 3 |
|---|---|
| | |

Do not write outside the box

**4** A supermarket sells bags of baking potatoes that are labelled as either '2 kg' or '5 kg'.

The weights of the contents of the bags labelled as '2 kg' are normally distributed with mean 2.20 kg and standard deviation 0.08 kg.

The weights of the contents of the bags labelled as '5 kg' are normally distributed with mean 5.26 kg and standard deviation 0.13 kg.

These weights are independent from bag to bag.

**(a)** Jo randomly selects one '2 kg' bag and one '5 kg' bag of potatoes.

**(i)** Find the probability that **both** of these bags have contents that weigh more than their respective labelled weights.

**[5 marks]**

**(ii)** Find the mean and the standard deviation of the **total** weight of the contents of the two bags that Jo has selected.

**[2 marks]**

**(b)** Heather randomly selects five '2 kg' bags and two '5 kg' bags of potatoes.

Find the probability that the total weight of the contents of her five '2 kg' bags exceeds the total weight of the contents of her two '5 kg' bags.

**[5 marks]**

**(c)** The supermarket also sells baking potatoes individually so that customers can select their own. These are priced at £0.90 per kilogram.

The weights for these individual potatoes are normally distributed with mean 0.25 kg and standard deviation 0.015 kg.

Arjun requires 8 baking potatoes. He had intended to buy a '2 kg' bag containing 8 potatoes that was priced at £1.85. However, he now thinks it would be cheaper to buy them individually.

Arjun selects 8 of these individual baking potatoes. Assuming that this may be regarded as a random selection:

**(i)** write down the probability that the total weight of these 8 potatoes is more than 2 kg ;

**[1 mark]**

**(ii)** find the probability that it is cheaper for Arjun to buy these 8 potatoes rather than a '2 kg' bag.

**[5 marks]**

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 4 |
|---|---|
| | |

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 4 |
|---|---|
| | |

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 4 |
|---|---|
| | |

**Turn over ▶**

Do not write outside the box

**5** Peak Expiratory Flow Rate (PEFR) is used to check lung function by using a PEFR meter to measure how fast a person can expel air from their lungs.

**(a)** After completing a standard set of exercises, a male with severe asthma will, on average, experience a reduction in PEFR of $30$ litres per minute.

A random sample of $9$ males with severe asthma had their PEFR measured before and after a walk on a cold winter's day. The reductions in PEFR values (in litres per minute) for these $9$ males were as follows.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 12 | 41 | 108 | 76 | 76 | 2 | 63 | 72 | 59 |

Assume that reductions in PEFR values can be modelled by a normal distribution.

Test, at the $1\%$ level of significance, whether a walk on a cold winter's day produces a greater mean reduction in PEFR values for males with severe asthma than $30$ litres per minute.

**[7 marks]**

**(b)** Using data from a different random sample of males with severe asthma, the same test as that in part **(a)** was carried out.

State, in context, the conclusion that must have been made from this test if it was subsequently found that a Type $\text{II}$ error had been made.

**[2 marks]**

| 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

**6 (a)** At a recent referendum, Brideston Village Hall was used as a polling station.

The number of voters arriving at the hall during a $10$-minute interval between $9$ am and $12$ noon on the morning of the referendum can be modelled by a Poisson distribution with mean $6.5$ .

**(i)** Find the probability that during a particular $10$-minute interval between $9$ am and $12$ noon the number of voters arriving at the hall was $5$ or fewer.

**[1 mark]**

**(ii)** Use a distributional approximation to find the probability that the number of voters arriving at the hall during a particular **hour** between $9$ am and $12$ noon was greater than $50$.

**[6 marks]**

**(b)** The number of voters arriving at the hall between $6$ pm and $9$ pm in the evening can also be modelled by a Poisson distribution. However, the evening arrival rate might differ from the morning arrival rate.

During the hour from $7$ pm to $8$ pm, $66$ voters arrived.

**(i)** Calculate an approximate $95\%$ confidence interval for the number of voters arriving at the hall **per hour** during the evening.

**[3 marks]**

**(ii)** Use your confidence interval to compare the morning arrival rate of voters at Brideston Village Hall to the evening arrival rate.

**[3 marks]**

**(iii)** Give **two** reasons why your confidence interval is approximate rather than exact.

**[2 marks]**

**(c)** The neighbouring Coniford Village Hall was also used as a polling station for the referendum.

During a randomly selected $10$-minute interval between $9$ am and $12$ noon, $2$ voters arrived at Coniford Village Hall.

Test, at the $10\%$ level of significance, whether the morning arrival rate at Coniford Village Hall differs from that at Brideston Village Hall.

**[5 marks]**

| QUESTION PART REFERENCE | **Answer space for Question 6** |
|---|---|
| | |
| | |
| | |

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 6 |
|---|---|
| | |

**Turn over ▶**

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 6 |
|---|---|
| | |

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 6 |
|---|---|
| | |

**Turn over ▶**

Do not write outside the box

| QUESTION PART REFERENCE | Answer space for Question 6 |
|---|---|
| | |

**END OF QUESTIONS**

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**

**Copyright information**

For confidentiality purposes acknowledgements of third party copyright material are published in a separate booklet which is available for free download from www.aqa.org.uk after the live examination series.

Copyright © 2018 AQA and its licensors. All rights reserved.