![cea logo]
Rewarding Learning

General Certificate of Secondary Education
2018

Statistics

Unit 1
Higher Tier

[GST12]

THURSDAY 21 JUNE, MORNING

MARK
SCHEME

GCSE Stastics

Introduction

The mark scheme normally provides the most popular solution to each question. Other solutions given by candidates are evaluated and credit given as appropriate; these alternative methods are not usually illustrated in the published mark scheme.

The marks awarded for each question are shown in the right hand column and they are prefixed by the letters **M**, **A** and **MA** as appropriate. The key to the mark scheme is given below:

M   indicates marks for correct method.

A   indicates marks for accurate working, whether in calculation, reading from tables, graphs or answers.

MA   indicates marks for combined method and accurate working.

A later part of a question may require a candidate to use an answer obtained from an earlier part of the same question. A candidate who gets the wrong answer to the earlier part and goes on to the later part is naturally unaware that the wrong data is being used and is actually undertaking the solution of a parallel problem from the point at which the error occurred. If such a candidate continues to apply correct method, then the candidate's individual working must be **followed through** from the error. If no further errors are made, then the candidate is penalised only for the initial error. Solutions containing two or more working or transcription errors are treated in the same way. This process is usually referred to as "follow-through marking" and allows a candidate to gain credit for that part of a solution which follows a working or transcription error.

It should be noted that where an error trivialises a question, or changes the nature of the skills being tested, then as a general rule, it would be the case that not more than half the marks for that question or part of that question would be awarded; in some cases the error may be such that no marks would be awarded.

Positive marking:

It is our intention to reward candidates for any demonstration of relevant knowledge, skills or understanding. For this reason we adopt a policy of **following through** their answers, that is, having penalised a candidate for an error, we mark the succeeding parts of the question using the candidate's value or answers and award marks accordingly.

Some common examples of this occur in the following cases:

(a)   a numerical error in one entry in a table of values might lead to several answers being incorrect, but these might not be essentially separate errors;

(b)   readings taken from candidates' inaccurate graphs may not agree with the answers expected but might be consistent with the graphs drawn.

When the candidate misreads a question in such a way as to make the question easier, only a proportion of the marks will be available (based on the professional judgement of the examiner).

1 (a) 21 315 A1

(b) 461 A1

(c) Suitable valid reason, e.g.
- Some drivers may have had more than one offence recorded
- Headline does not include time frame
- Not all offences relate to driving A1 3

2 (a) Valid example of closed question with suitable mutually exclusive, exhaustive response boxes. A2

(b) Appropriate advantage, e.g.
- Consistency of responses
- No interpretation required by researcher
- Easy to analyse
- Allows comparability A1

Appropriate disadvantage, e.g.
- Assumes respondent has understood the question
- Response section may not include the exact answer a respondent wishes to give
- Can force responses A1

(c) She should have conducted a pilot survey to identify problems with the questions. A2 6

3 Median/average has decreased so boot camp was generally successful. MA1 A1

One additional comment, e.g.
- Range/IQR decreased
- Improved consistency of RHR among the group A1 3

4 **(a)** Any advantage, e.g.
- no need to have a row for every possible time A1

Any disadvantage, e.g.
- accuracy is lost by grouping A1

(b)

Time	Frequency, f	t	ft
$0 < t \leqslant 10$	343	5	1715
$10 < t \leqslant 20$	214	15	3210
$20 < t \leqslant 30$	197	25	4925
$30 < t \leqslant 40$	145	35	5075
$40 < t \leqslant 50$	72	45	3240
$50 < t \leqslant 60$	29	55	1595
TOTAL	**1000**		**19 760**

MA2

$$\bar{t} = \frac{19\,760}{1000}$$

M1

= 19.76 seconds A1

(c) Claim is justified because estimated mean is less than 20 A2

(d) (i) No A1

(ii) Any unsatisfied property, e.g.
- distribution not symmetrical
- histogram has a positive skew A1   10

5 **(a)** The people who live in the town. A1

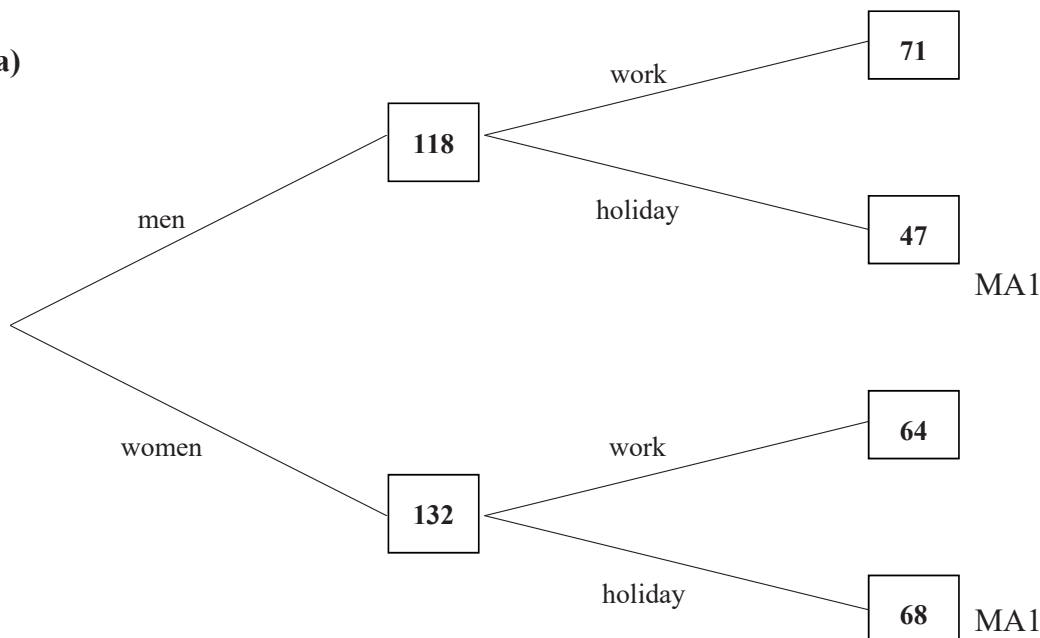**(b)** 'Recently' not specific enough/open to interpretation A1

No box for not visiting the library at all A1   3

6 (a)

work → **71**

118

holiday → **47**   MA1

men

women

132   work → **64**

holiday → **68**   MA1

(b) $P(H \mid F) = \dfrac{68}{132} \left(= \dfrac{17}{33}\right)$ MA2

(c) $P(M \mid H) = \dfrac{47}{47 + 68} = \dfrac{47}{115}$ MA1 A1   6

7 (a) Mean $= \dfrac{3 + 8 + 9 + 5 + 12 + 4 + 2 + 1}{8}$ M1

$= 5.5$ times   A1

(b) All the numbers have the same frequency so there is no mode   A1

(c) Decrease   A1

(d) People living closer to the cinema go to it more often   A1

(e) Quantitative   A1

Discrete   A1

(f) Negative correlation, so people living closer to the cinema go more often.   A2   9

8 **(a)** This would be very expensive/time consuming. A1

(b) They allow the production manager to monitor whether the process complies with pre-set standards/remains under control. A1

(c) Warning lines drawn at 413.8 and 416.2 MA1

Action lines drawn at 413.2 and 416.8 MA1

Lines labelled correctly. MA1

(d) The mean of sample 6 was below the lower warning line so the manager should have been alerted to a potential problem with the process/taken another sample immediately. A2

The mean of sample 7 is above the upper action line so the process should have been stopped/machine reset or repaired. A2   9
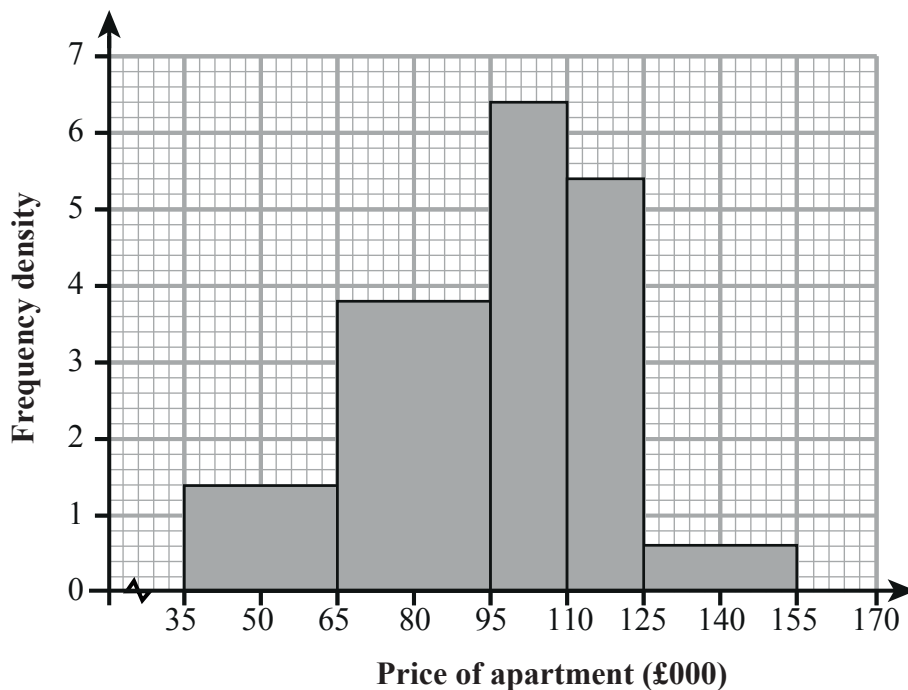
9 **(a)**

Price (£000)	No. of Apartments	Frequency density	
$35 \leqslant p < 65$	42	1.4	
$65 \leqslant p < 95$	114	3.8	
$95 \leqslant p < 110$	96	6.4	M1
$110 \leqslant p < 125$	81	5.4	
$125 \leqslant p \leqslant 155$	18	0.6	MA1

MA2

(b) Any two suitable reasons, e.g.
- results are for a sample so not all apartments are represented
- no information on sampling method is given so sample may not faithfully represent the population
- there is no way of telling exactly which values are contained in the range $125 \leqslant p \leqslant 155$

A2   6

10 (a) Raw data exists/occurs in cycles of 4 as 4 quarters in a year   A1

(b) $\dfrac{130 + 115 + 97 + 120}{4} = 115.5$ M1

$\dfrac{115 + 97 + 120 + 134}{4} = 116.5$

$\dfrac{97 + 120 + 134 + 125}{4} = 119$ A2

(c) Points correctly plotted and trend line drawn   MA2

(d) The number of complaints is increasing over time   A1

(e) $4 \times 121 - (120 + 134 + 125)$ MA2

$= 105$

or

$4 \times 122 - (120 + 134 + 125)$

$= 109$ A1   10

11 (a) (6, 28) indicated   MA1

(b) The expected score for a candidate who took no lessons.   A2

(c) $y = 50.6 + 1.58 \times 11$ M1

$= 67.98$

$= 68$ A1

(d) $r_s = 1 - \left(\dfrac{6 \times 4}{8(8^2 - 1)} \right)$ M1MA1

$= 0.952$ A1

(e) Anna is unlikely to be correct since the value of r_s indicates a strong positive correlation between the judges' scores.   A2   10

12 (a) $1.01 \times 3.09 = £3.12$ MA1 A1

(b) $\dfrac{3.17}{3.22} \times 100$ MA2

    $= 98.4$ A1

(c) The price of coffee is 98.4% of what it was in 2016   A2

(d) $0.6 \times £340\,540 \times 1.06$ M1 MA2

    $£216\,583(.44)$ A1

(e) $\dfrac{(106 \times 60) + (103 \times 25) + (107 \times 15)}{60 + 25 + 15}$ M1 MA2

    $= 105.4$ A1   15

13 (a) $z = \dfrac{35 - 33.4}{1.7}$ M1 MA1

    $= 0.941$ A1

(b) $-1.455 = \dfrac{x - 31.6}{1.1}$ MA1

    $x = 30\,\text{mm}$ A1

(c) $z_1 = \dfrac{32 - 33.4}{1.7}$ M1

    $= -0.8235\ldots$ A1

    $z_2 = \dfrac{32 - 31.6}{1.1}$

    $= 0.3636\ldots$ MA1

    Stone is more likely to have come from the second location   A1

(d) The distributions must be normal.   A1   10

          Total   **100**