AQA

A-LEVEL
Statistics
SS02
Mark scheme

6380
June 2018

Version/Stage: 1.0 Final

Mark schemes are prepared by the Lead Assessment Writer and considered, together with the relevant questions, by a panel of subject teachers. This mark scheme includes any amendments made at the standardisation events which all associates participate in and is the scheme which was used by them in this examination. The standardisation process ensures that the mark scheme covers the students' responses to questions and that every associate understands and applies it in the same correct way. As preparation for standardisation each associate analyses a number of students' scripts. Alternative answers not already covered by the mark scheme are discussed and legislated for. If, after the standardisation process, associates encounter unusual answers which have not been raised they are required to refer these to the Lead Assessment Writer.

It must be stressed that a mark scheme is a working document, in many cases further developed and expanded on the basis of students' reactions to a particular paper. Assumptions about future mark schemes on the basis of one year's document should be avoided; whilst the guiding principles of assessment remain constant, details will change, depending on the content of a particular examination paper.

Further copies of this mark scheme are available from aqa.org.uk

Key to mark scheme abbreviations

M	mark is for method
m or dM	mark is dependent on one or more M marks and is for method
A	mark is dependent on M or m marks and is for accuracy
B	mark is independent of M or m marks and is for method and accuracy
E	mark is for explanation
√or ft or F	follow through from previous incorrect result
CAO	correct answer only
CSO	correct solution only
AWFW	anything which falls within
AWRT	anything which rounds to
ACF	any correct form
AG	answer given
SC	special case
OE	or equivalent
A2,1	2 or 1 (or 0) accuracy marks
–*x* EE	deduct *x* marks for each error
NMS	no method shown
PI	possibly implied
SCA	substantially correct approach
c	candidate
sf	significant figure(s)
dp	decimal place(s)

No Method Shown

Where the question specifically requires a particular method to be used, we must usually see evidence of use of this method for any marks to be awarded.

Where the answer can be reasonably obtained without showing working and it is very unlikely that the correct answer can be obtained by using an incorrect method, we must award **full marks**. However, the obvious penalty to candidates showing no working is that incorrect answers, however close, earn **no marks**.

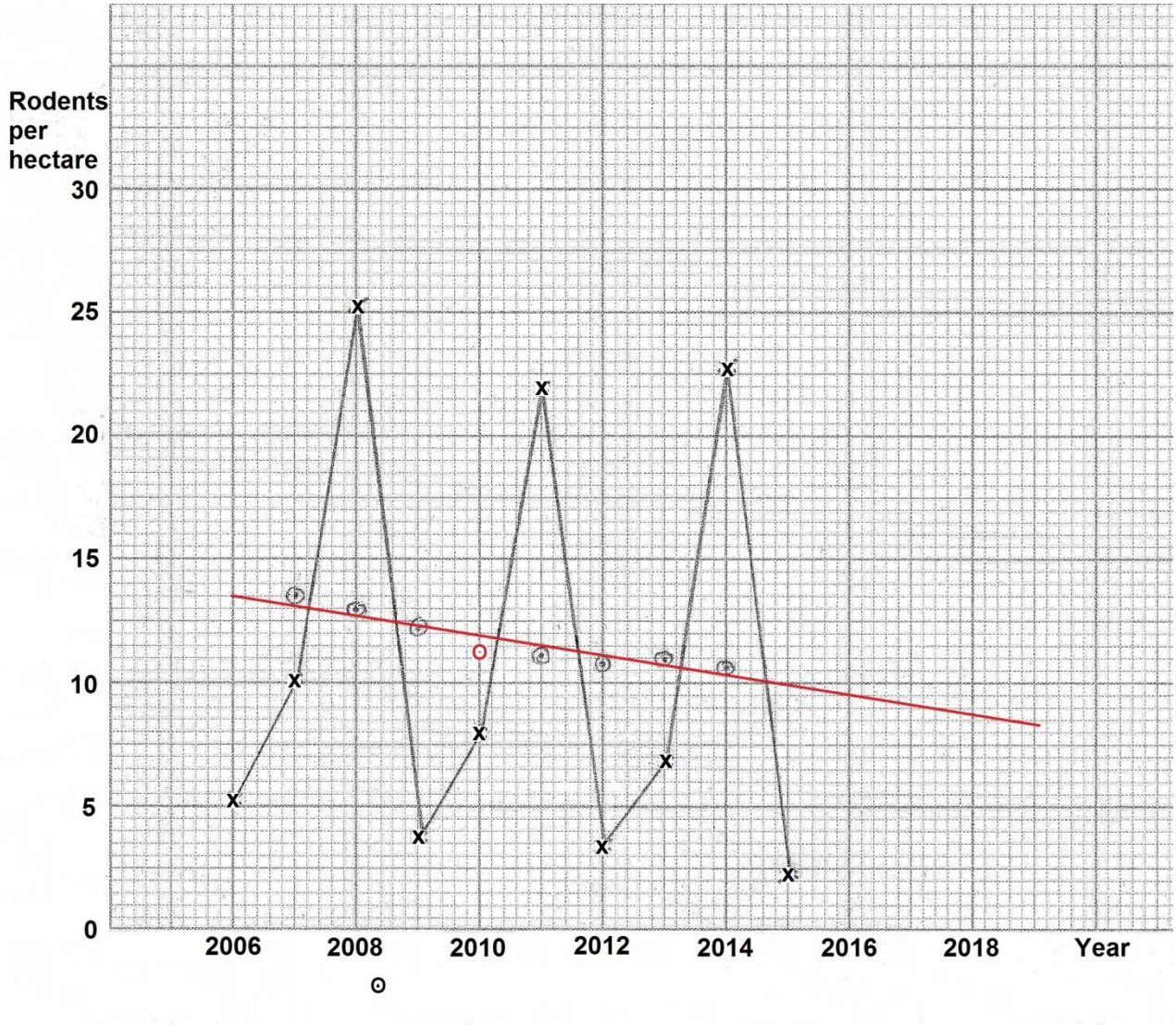
Where a question asks the candidate to state or write down a result, no method need be shown for full marks.

Where the permitted calculator has functions which reasonably allow the solution of the question directly, the correct answer without working earns **full marks**, unless it is given to less than the degree of accuracy accepted in the mark scheme, when it gains **no marks**.

Otherwise we require evidence of a correct method for any marks to be awarded.

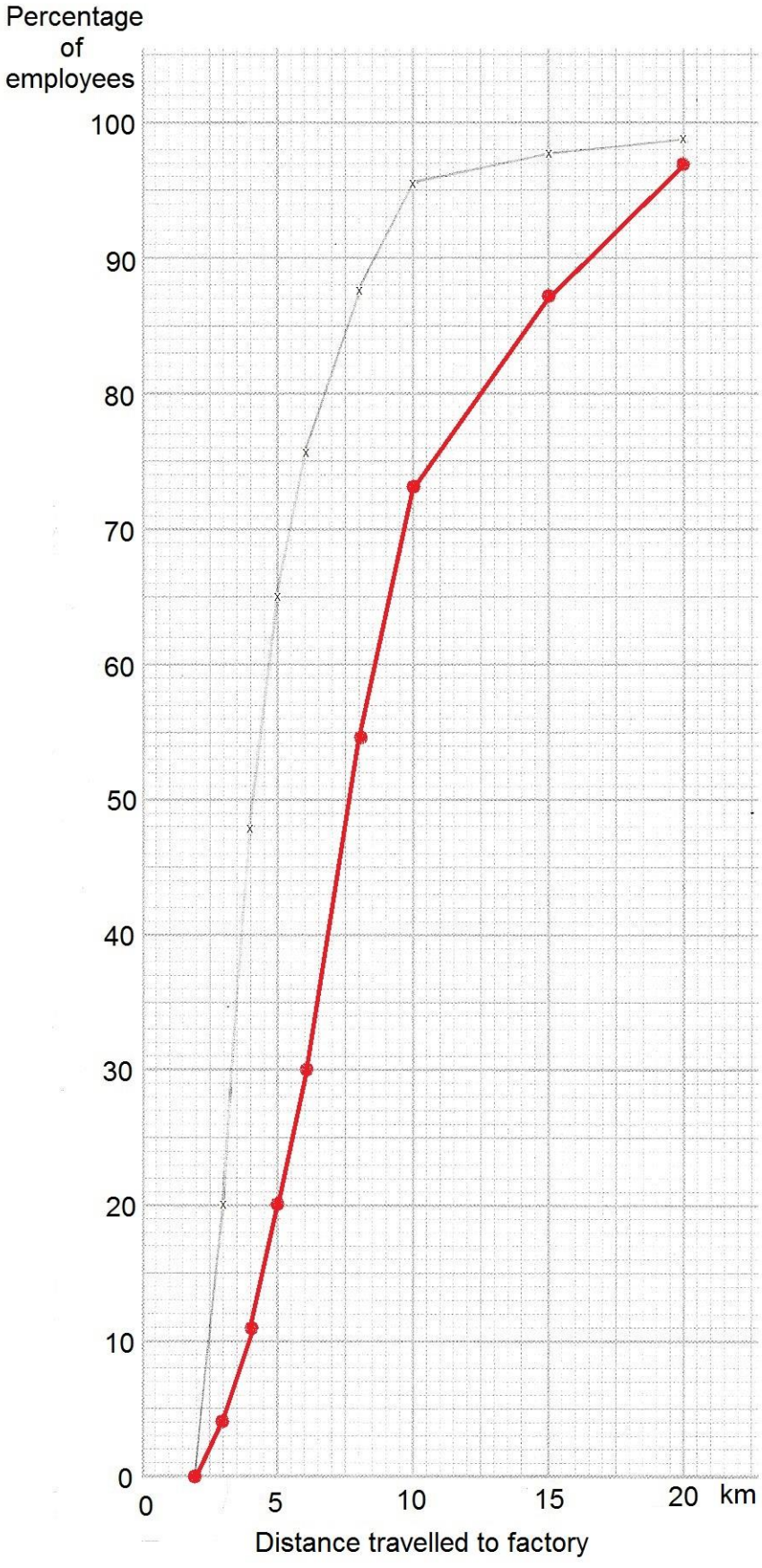
Q	Solution	Marks	Total	Comments
1 (a)	0.4, 40% or $^2/_5$	B1	**1**	Any form but not simply 40
(b)	0.6×0.25 $\times 2 = 0.3$ (or 30% or $^3/_{10}$)	M1 A1	**2**	If unsupported answer 0.15 is given allow this M1 Any equivalent form
(c)	Mean $= 1 \times 0.24 + 2 \times 0.36 \ldots + 5 \times 0.03$ $= 2.44$ $1^2 \times 0.24 + 2^2 \times 0.36 \ldots + 5^2 \times 0.03 - 2.44^2$ $= 0.24 + 1.44 + 1.35 + 3.52 + 0.75 - 2.44^2$ $= 7.3 - 2.44^2 = 1.3464$ and $\sqrt{(7.3 - 2.44^2)} = \sqrt{1.3464}$ (= 1.16)	M1 A1 M1 A1 A1	**5**	 CAO. Unsupported 2.44 scores B2 **Complete** method (their 2.44) Products or results of products Either of these AG
(d)	E(C) = 68 + 24 × 1.44 = £102.56 $\sigma_{c} = 24 \times 1.16 = 27.8(4)$ May use new table	B1 B1	**2**	102.56 or 102.6 or 103 27.8, or AWFW 27.80 to 27.90
		Total	**10**	

Q	Solution	Marks	Total	Comments
2 (a)(i)	(3.7 + 8.0 + 21.9)/3	M1		Any three consecutive values from table added and then divided by 3
	= 11.2	A1	**2**	CAO Value must be stated
(a)(ii)	Accurate plot	B1		Within half a square
	Reasonable trend line	B1	**2**	Points above and below
(a)(iii)	Seasonal variation (with a 3-year cycle)	B1	**1**	Or detailed description of cycle pattern
(b)	Seasonal variation = (12 + 11 + 12) ÷ 3	M1		3 values (±1) added and ÷ 3 (may be cm)
	= 11.7 (cm gives (4.8 + 4.4 + 4.8) ÷ 3 = 4.7)	A1		AWFW 11.5 to 12.0 (cm 4.6 to 4.8)
	Trend value = 9.0	B1		AWFW 8.5 to 9.5 (cm just using graph)
	Estimate = 20.8	A1	**4**	AWFW 20 to 21.5
(c)	2016 seems to follow the previous pattern	E1		OE
	2017 well below estimate – the pattern has changed	E1	**2**	OE
		Total	**11**	



Q	Solution	Marks	Total	Comments
3	$P(\leq 1) = 0.199(1)$	M1		Stated or implied
(a)	$1 - 0.199 = 0.801$	A1	**2**	AWRT
(b)	In 3 years Steve expects 12 migraines	M1		PI
	$P(\leq 12) - P(\leq 11)$	M1		Or formula or calculator. PI by correct answer.
	$= 0.114(4)$	A1	**3**	AWRT 0.114
(c)	Use of Po(14)	M1		AWRT 0.109, 0.176, 0.260, 0.740, 0.824, or 0.891 seen
	$1 - 0.1757 = 0.824(3)$	A1	**2**	AWRT 0.824
(d)	$P(R = 0) \times P(S > 0) + P(R = 1) \times P(S > 1)$	M1		Complete method
	$= 0.0498 \times (1 - 0.0183) + 0.1493 \times (1 - 0.0916)$	A1		Either product correct
	$= 0.184(5)$	A1	**3**	AWFW 0.184 to 0.185
		Total	**10**	

Q	Solution	Marks	Total	Comments
4 (a)(i)	<table><tr><th></th><th>Cumulative percentage</th></tr><tr><td>$x \leq 2$</td><td>0</td></tr><tr><td>$2 < x \leq 3$</td><td>4</td></tr><tr><td>$3 < x \leq 4$</td><td>10.5</td></tr><tr><td>$4 < x \leq 5$</td><td>20</td></tr><tr><td>$5 < x \leq 6$</td><td>30</td></tr><tr><td>$6 < x \leq 8$</td><td>54.5</td></tr><tr><td>$8 < x \leq 10$</td><td>73</td></tr><tr><td>$10 < x \leq 15$</td><td>87</td></tr><tr><td>$15 < x \leq 20$</td><td>97</td></tr><tr><td>$x > 20$</td><td></td></tr></table> Accurately plotted and joined	M1 M1 A1 A1	**4**	Attempt to accumulate (16, 42, 80, 120, 218, 292, 348, 388) Attempt to change to percentage (4, 6.5, 9.5, 10, 24.5, 18.5, 14, 10) Accurate results, all correct. Within one small square. Condone anything after $x = 20$ as long as it lies between 97 and 100 Allow full marks from graph alone01
(a)(ii)	8.5	B1	**1**	AWFW 8.3 to 8.7
(b)	The greatest distance travelled by an employee (and/or the shortest distance)	E1	**1**	OE
(c)(i)	Part-time median = 4.1 Full-time median = 7.7 (Or difference of 3.6) Full-time employees travel on average (3.6 km) more than part-time employees	B1 E1dep		4.0 to 4.2 7.5 to 7.9 B1 for both or difference 3.3 to 3.9 Dep on B1. Right way round.
(c)(ii)	Part-time IQR = 6 – 3.1 = 2.9 Full-time IQR = 10.7 – 5.6 = 5.1 (Or difference of 2.2) Distances for full-time employees more spread out than for part-time employees (Or IQR around 2.2 greater)	B1 E1dep	**4**	2.7 to 3.0 4.8 to 5.4 B1 for both or difference 1.8 to 2.7 Dep on B1. Right way round.
		Total	**10**	


Percentage of employees
100
90
80
70
60
50
40
30
20
10
0
0
5
10
15
20
km
Distance travelled to factory

Q	Solution	Marks	Total	Comments
5 (a)	(Rounding errors) because the data are given to the nearest 1000	E1	**1**	Accept simply "rounding errors"
(b)(i)	105 (000) ÷ 17709 = 5.9(3)	M1 A1	**2**	Allow without 000 AWRT 5.9
(ii)	7904 (000 000) ÷ 33(000) ($)240 000	M1 A1	**2**	Allow without 000 000 and 000 AWRT
(c)	11(.0)%	B1	**1**	11 or AWRT 11.0
(d)	$\frac{3653 - 3521}{3653} \times 100$ $\frac{11556 - 11488}{11556} \times 100$, $\frac{6736 - 6713}{6736} \times 100$ 3.61% > 0.59% and 0.34% **Alternative** All 3 differences calculated 132, 68 and 23 Pointed out that mining has the biggest difference **and** the smallest 2015 GDP So % difference for mining is greatest	M1 M1 A1 (M1) (M1) (A1)	**3**	Both 1 d.p. or better All three must be seen No further calculation necessary No further calculation necessary
		Total	**9**	

Q	Solution	Marks	Total	Comments
6	The sample is a random sample	E1		
(a)	from a normal population of weights	E1		Context. Not "a normal sample"
	H_0: $\mu = 405.3$			
	H_1: $\mu \neq 405.3$	B1		For both
	Critical z value = ±1.96	B1		AWRT
	$\bar{x} = 386.6$	B1		CAO
	Test statistic = ('386.6' – 405.3)÷(31.2/√5)	M1		Complete formula (ignore sign) (their 386.6)
	= –1.34	A1		AWRT
	–1.96 < –1.34 < 1.96			
	So accept H_0, **no significant evidence** that mean weight (of puffins) has changed	A1dep	**8**	Dep on A1 and critical value B1 Must include "mean weight" Not absolute "weight has not changed"
(b)	Test statistic = (400.9 – 405.3)÷(28.4/√196) = (–)2.17	B1		AWRT ignore sign
	CV = 2.054 for 2% significance and 2.326 for 1% significance	M1		For either to at least 3sf
	So least possible α is 2	A1		CAO
	Alternative			
	p-value calculated as 0.015(0)	(B1)		AWRT 0.015
	0.01 < 0.015 < 0.02	(M1)		PI by correct final choice
	So least possible α is 2	(A1)	**3**	CAO
(c)	Type II error	B1	**1**	
		Total	**12**	

02

6 (a)	Alternative (1) for last three marks			
	Critical values at $405.3 \pm 1.96 \times (31.2/\sqrt{5})$	M1		Complete formula
	$= 380.0$ & 432.6	A1		AWRT 380 and 433
	$380.0 < 386.6 < 432.6$ So accept H_0, **no significant evidence** that mean weight (of puffins) has changed	A1dep	**8**	Dep on A1 and critical value B1 Must include "mean weight" Not absolute "weight has not changed"

6 (a)	Alternative (2) for last three marks			
	95% conf interval at $386.6 \pm 1.96 \times (31.2/\sqrt{5})$	M1		Complete formula
	$= 359.3$ & 413.9	A1		AWRT 359 and 414
	$359.3 < 405.3 < 413.9$ So accept H_0, **no significant evidence** that mean weight (of puffins) has changed	A1dep	**8**	Dep on A1 and critical value B1 Must include "mean weight" Not absolute "weight has not changed"

6 (a)	Alternative (3) for last five marks			
	$\bar{x} = 386.6$	B1		CAO
	z value for this is ('386.6' – 405.3)÷$(31.2/\sqrt{5})$	M1		Complete formula. PI by p value
	$= -1.34$	A1		AWRT PI by correct p value
	Giving a p value of 0.090 (or 0.180)			
	$0.090 > 0.025$ (or $0.180 > 0.05$)	A1		Must be correct comparison AWRT 0.090 or 0.180
	So accept H_0, **no significant evidence** that mean weight (of puffins) has changed	A1dep	**8**	Dep on both A1s Must include "mean weight" Not absolute "weight has not changed"

Q	Solution	Marks	Total	Comments
7 (a)	The list does not tell him the sex or age of the participants ….	E1		The problem with the list
	so he cannot pick a sample in the right proportions	E1	**2**	The consequence for the sample
	Alternative for one mark The list does not give any information about whether the runner exercises regularly	(E1)		
(b)(i)	Using random number tables or function select 4-digit random number	B1		Must specify 4-digit
	ignoring (0000 and) >2000 and repeats	B1		Condone not mentioning 0000
	continue until 40 have been selected (and measure those runners)	B1		
(b)(ii)	eg. difficult to find the numbered people at the end of the run some of the numbers chosen for his sample may belong to people who do not finish the run.	E1	**4**	Must be in this context and practical, not statistical
(c)(i)	Randomly choose one of the first 50 to finish	B1		Must specify random
	Choose every 50^{th} person (after that) to be in the sample	B1		Indep of previous B1
(c)(ii)	Advantage: Eg. Should space out the people who need to be measured. Should give a spread of fitness levels.	E1		Must be in this context
	Disadvantage: Eg. May get a large number crossing the line close together – hard to manage	E1	**4**	Must be in this context
(d)(i)	Cluster sampling	B1		
(d)(ii)	Advantage: Eg. Easy to spot who is in the sample. Can give them instructions before the run	E1		Must be in this context
	Disadvantage: Eg. May be very unrepresentative, as one charity may be all the same age, all the same sex, all the same1 fitness level.	E1	**3**	Must be in this context
		Total	**13**	