Teacher Support Materials 2009

Statistics GCE

Paper Reference SS02

The Assessment and Qualifications Alliance (AQA) is a company limited by guarantee registered in England and Wales (company number 3644723) and a registered charity (registered charity number 1073334). Registered address: AQA, Devas Street, Manchester M15 6EX.
Dr Michael Cresswell, Director General.

Question 1

A charity shop opens for 8 hours each day. The daily number of customers requesting a refund for an earlier purchase may be modelled by a Poisson distribution with mean 0.6.

(a) Find the probability that on a particular day there are:

(i) no requests for refunds;

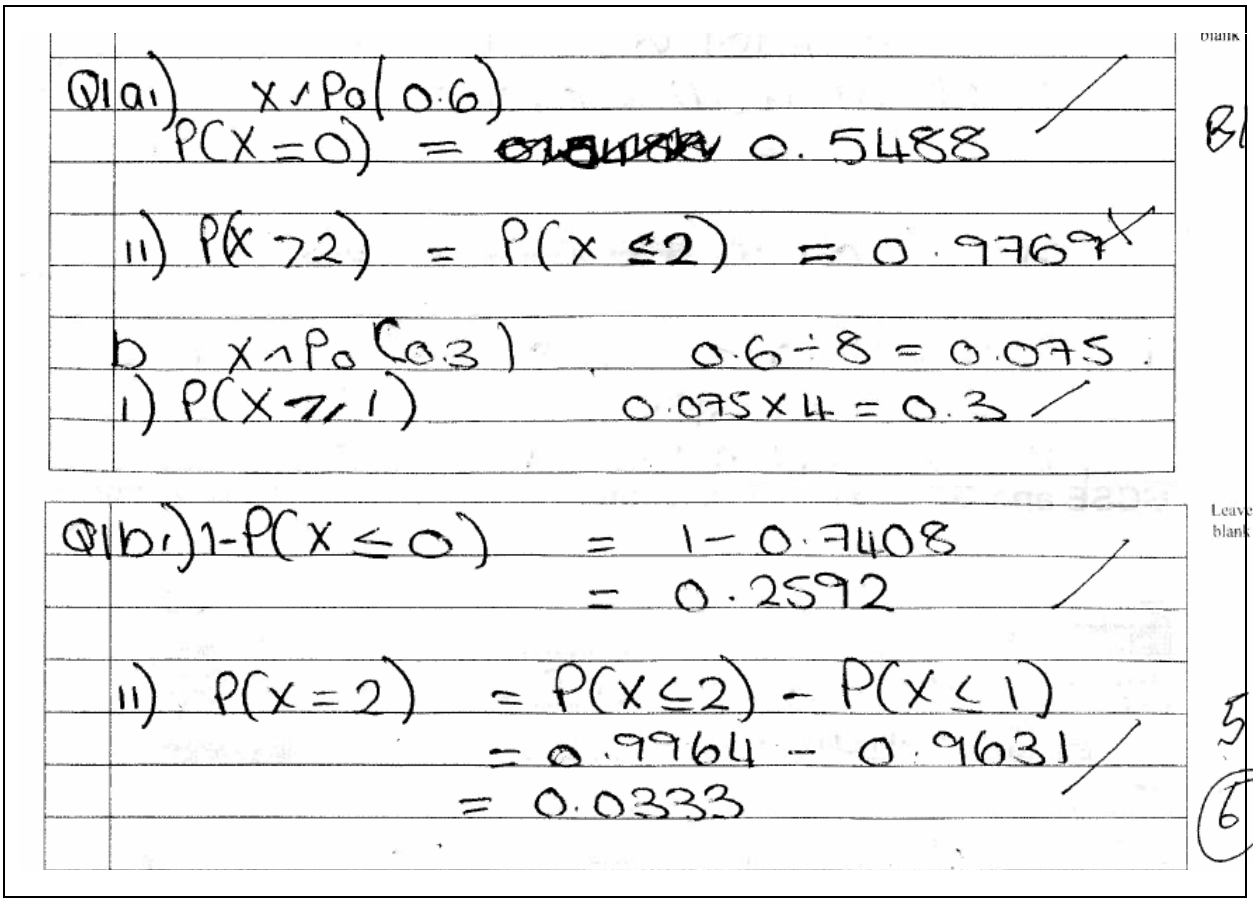
(ii) more than 2 requests for refunds. *(3 marks)*

(b) On a particular day, the shop manager is absent for 4 hours and leaves an assistant in charge of the shop. Find the probability that, during the manager's 4-hour absence, there are:

(i) one or more requests for refunds;

(ii) exactly 2 requests for refunds. *(5 marks)*

Student Response



Commentary

Many candidates gained full marks for this question. Unfortunately this candidate spoilt a complete answer by failing to subtract her result in part (a)(ii) from 1.

Mark scheme

Q	Solution	Marks	Total	Comments
1(a)(i)	0.5488	B1		0.5485 ~ 0.549
(ii)	$P(X > 2) = 1 - P(X \leq 2)$	M1		
	$= 1 - 0.9769$			
	$= 0.0231$	A1	3	0.023 ~ 0.0232
(b)(i)	Po(0.3)	B1		Attempted use of Po(0.3)
(ii)	$P(X \geq 1) = 1 - P(X = 0)$	M1		
	$= 1 - 0.7408$			
	$= 0.259$	A1		0.259 ~ 0.26
(ii)	$P(X = 2) = 0.9964 - 0.9631$	M1		Method – includes Po(0.3)
	$= 0.0333$	A1	5	0.033 ~ 0.0334
	Total		**8**	

Question 2

Nazia owns an ice-cream van and sells tubs containing 1, 2, 3 or 4 scoops of ice cream. The prices, X pence, of these tubs together with the distribution of sales are shown in **Table 1**.

Table 1

	x	$P(X = x)$
1 scoop	50	0.40
2 scoops	95	0.16
3 scoops	135	0.24
4 scoops	170	0.20

(a) Find the mean and the standard deviation of X. *(4 marks)*

(b) Nazia is considering discontinuing the sale of tubs containing 1 scoop of ice cream. She estimates that if she does this:

- half of those customers who buy such tubs would instead buy tubs containing 2 scoops of ice cream;
- the other half would not buy any ice cream.

The new distribution of prices, Y pence, would be as shown in **Table 2**.

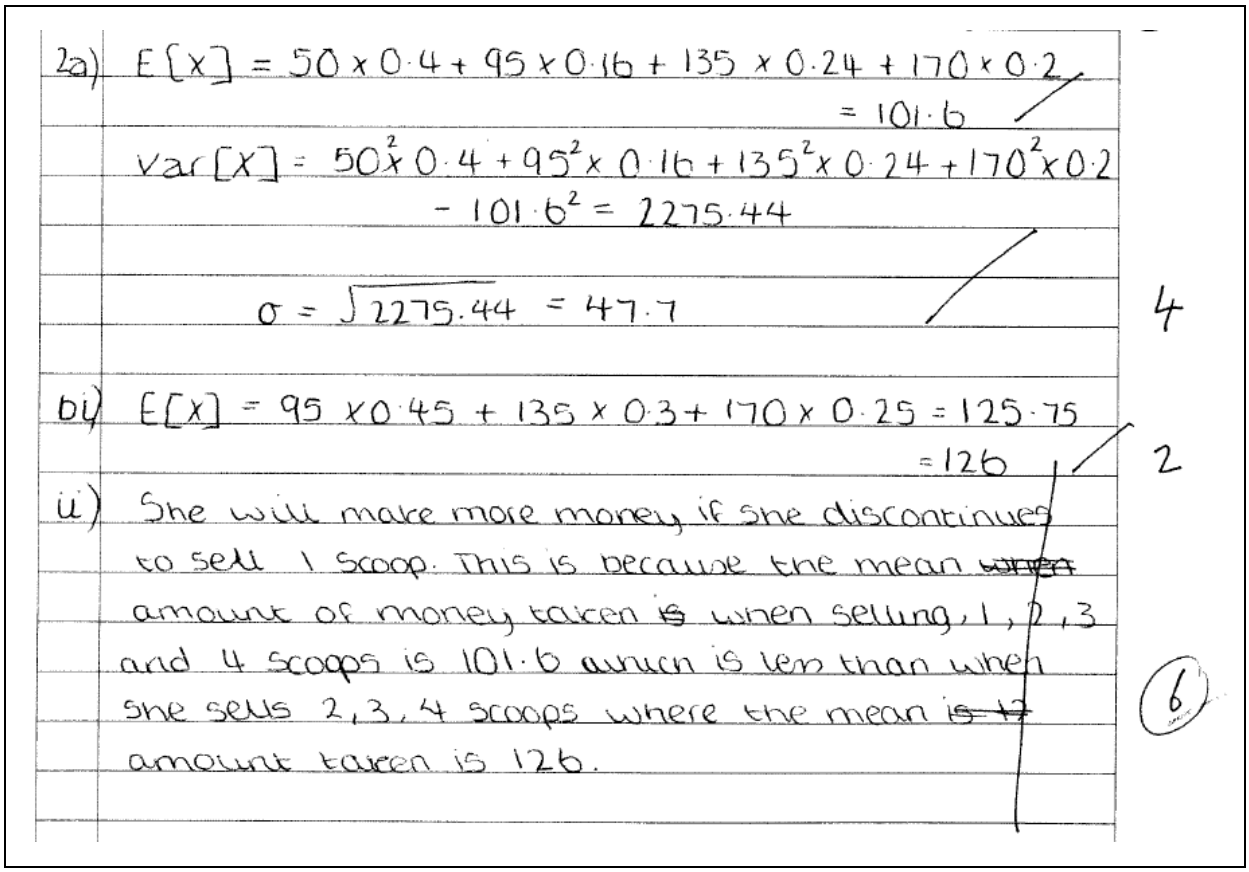
Table 2

	y	$P(Y = y)$
2 scoops	95	0.45
3 scoops	135	0.30
4 scoops	170	0.25

(i) Show that the mean value of Y is 126, correct to three significant figures. *(2 marks)*

(ii) By making a suitable calculation, advise Nazia as to whether, if her estimate is correct, she will increase the total amount of money taken if she discontinues selling tubs containing 1 scoop of ice cream. *(3 marks)*

Student Response



Commentary

Parts (a) and (b)(i) gained full marks. Part (a) could have been answered directly from a calculator, but the working was essential in part (b)(i). In part (b)(ii) the candidate compared the two means but took no account of the fact that Nazia would sell less tubs of ice-cream if she discontinued selling one-scoop tubs. This was a common error.

Mark Scheme

2(a)	$\mu = 101.6$	B2		CAO (allow 102) (or $E(X) = 50 \times 0.40 + 95 \times 0.16 + 135 \times 0.24 + 170 \times 0.20 = 101.6$ M1A1)
	$\sigma = 47.7$	B2	4	47.69 ~ 47.71 (or $E(X^2) = 50^2 \times 0.40 + 95^2 \times 0.16 + 135^2 \times 0.24 + 170^2 \times 0.20 = 12598$ $V(X) = 12598 - 101.6^2 = 2275.44$ s.d. $= \sqrt{2275.44} = 47.7$ M1A1)
(b)(i)	$E(X) = 95 \times 0.45 + 135 \times 0.30 + 170 \times 0.25 = 125.75$	M1		
	= 126 to 3sf	A1	2	CAO; AG
(ii)	Will lose 20% of customers $0.8 \times 125.75 = 100.6$	M1		Any relevant calculation attempted
	… which is less than 101.6.	m1		Valid comparison - their figures
	Hence, if estimate is correct, she will take less money.	A1	3	Correct conclusion based on correct working or lose $0.2 \times 50 = 10$; gain $0.2 \times (95 - 50) = 9$ or $100 \times 101.6 = 101.60$ $(100 - 120) \times 125.75 = 100.60$
	Total		**9**	

Question 3

3 **Table 3** shows the number of passenger journeys undertaken on local bus services in Great Britain each year from 1994/95 to 2004/05.

Table 3

Local bus services: passenger journeys: by area

Millions

	1994/95	1995/96	1996/97	1997/98	1998/99	1999/2000	2000/01	2001/02	2002/03	2003/04	2004/05
Great Britain	4402	4366	4333	4313	4231	4278	4319	4352	4444	4564	4609
London	1155	1193	1230	1281	1266	1294	1347	1422	1527	1692	1782
English Metropolitan Counties	1331	1292	1246	1232	1195	1178	1166	1154	1145	1114	1083
English Shire Counties	1271	1259	1260	1243	1242	1250	1247	1222	1210	1189	1167
All outside London	3247	3173	3103	3032	2965	2984	2972	2930	2917	2871	2828
England	3757	3744	3736	3755	3702	3722	3761	3798	3882	3995	4032
Scotland	513	494	467	438	413	442	443	449	452	457	465
Wales	132	127	130	120	116	114	116	104	110	111	113

Source: Department for Transport and *Annual Abstract of Statistics*, Office for National Statistics, 2006

(a) How many passenger journeys were undertaken on local bus services in Wales during 2004/05? *(2 marks)*

(b) (i) Describe the trend over the period 1994/95 to 2004/05 in the number of passenger journeys undertaken on local bus services in London.

(ii) Compare this trend with that on local bus services outside London over the same period. *(3 marks)*

(c) **Table 4** shows indices of fares for local bus services and also values of the Retail Prices Index in Great Britain each year from 1994/95 to 2004/05.

Table 4

Local bus services: fare indices: by area

Indices (1995 = 100)

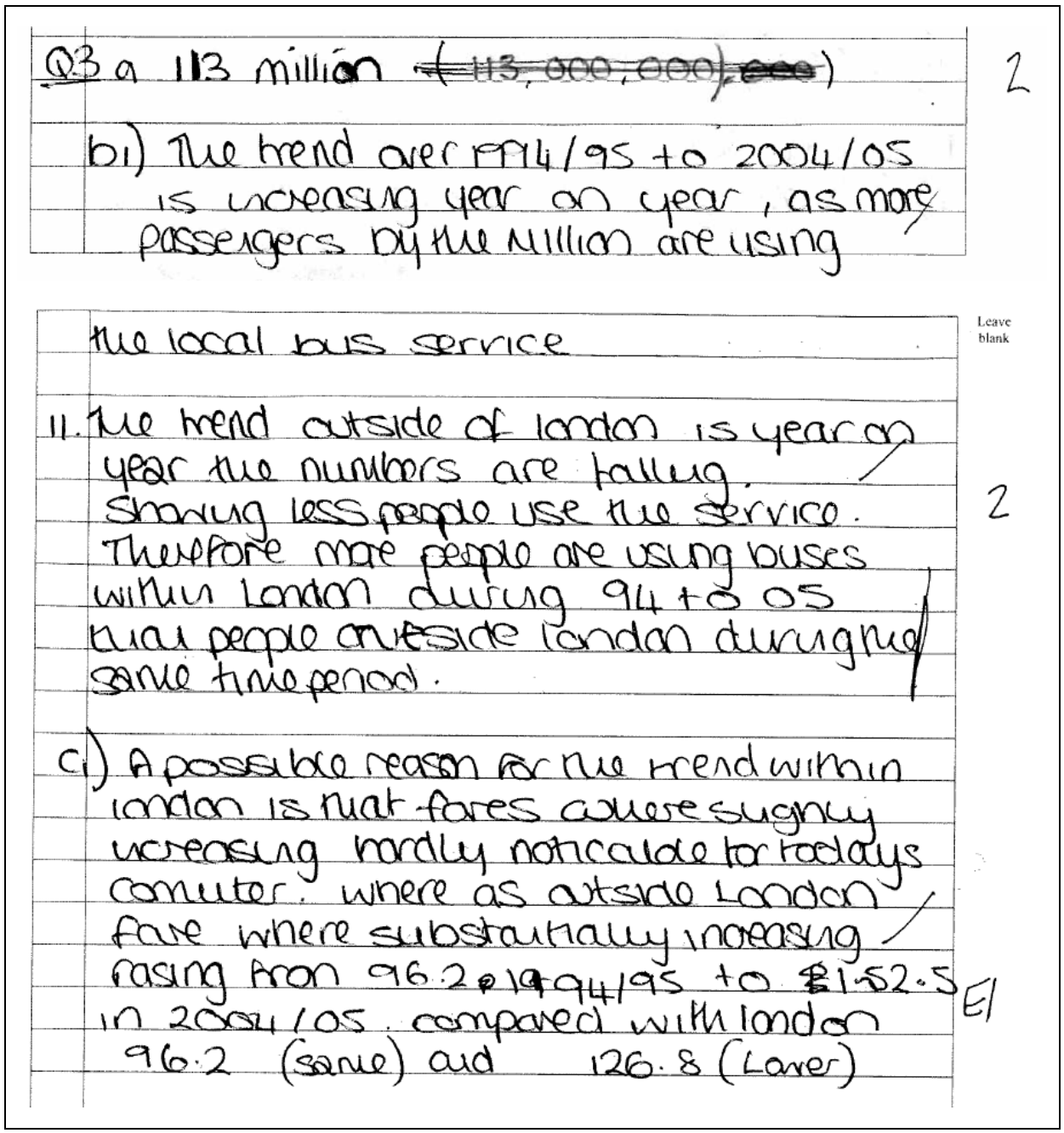
	1994/95	1995/96	1996/97	1997/98	1998/99	1999/2000	2000/01	2001/02	2002/03	2003/04	2004/05
Great Britain	96.7	101.2	106.3	112.0	117.1	122.0	126.4	130.6	134.5	139.1	145.7
London	96.2	101.1	105.4	109.3	113.7	117.2	117.2	115.5	114.8	116.9	126.8
English Metropolitan Counties	96.4	101.5	106.9	113.3	118.7	124.6	129.9	137.4	142.7	148.0	154.2
English Shire Counties	97.0	101.1	106.0	111.5	116.7	122.0	128.6	135.1	141.7	148.5	155.7
All outside London	96.8	101.2	106.6	112.8	118.2	123.4	129.2	135.3	140.8	146.3	152.5
England	96.7	101.2	106.1	111.4	116.5	121.5	125.9	130.3	134.2	139.1	146.2
Scotland	96.9	100.8	108.0	116.5	121.8	125.3	129.9	131.8	134.5	136.8	140.4
Wales	97.4	100.7	104.4	110.1	116.3	122.2	127.5	133.5	139.5	145.5	152.4
Retail Prices Index (1995 = 100)	97.5	100.7	103.1	106.5	109.9	111.6	114.9	116.6	119.1	122.4	126.2

Source: Department for Transport and *Annual Abstract of Statistics*, Office for National Statistics, 2006

(i) Give possible reasons for the trends considered in part (b), given the additional information in **Table 4**. *(3 marks)*

(ii) An index of motoring costs in Great Britain was 100 for the year 1994 and had risen to 122 by the year 2004. The corresponding figures for an index of rail fares in Great Britain were 100 and 136. Extend your answer to part (c)(i), given this further information. *(2 marks)*

Student Response



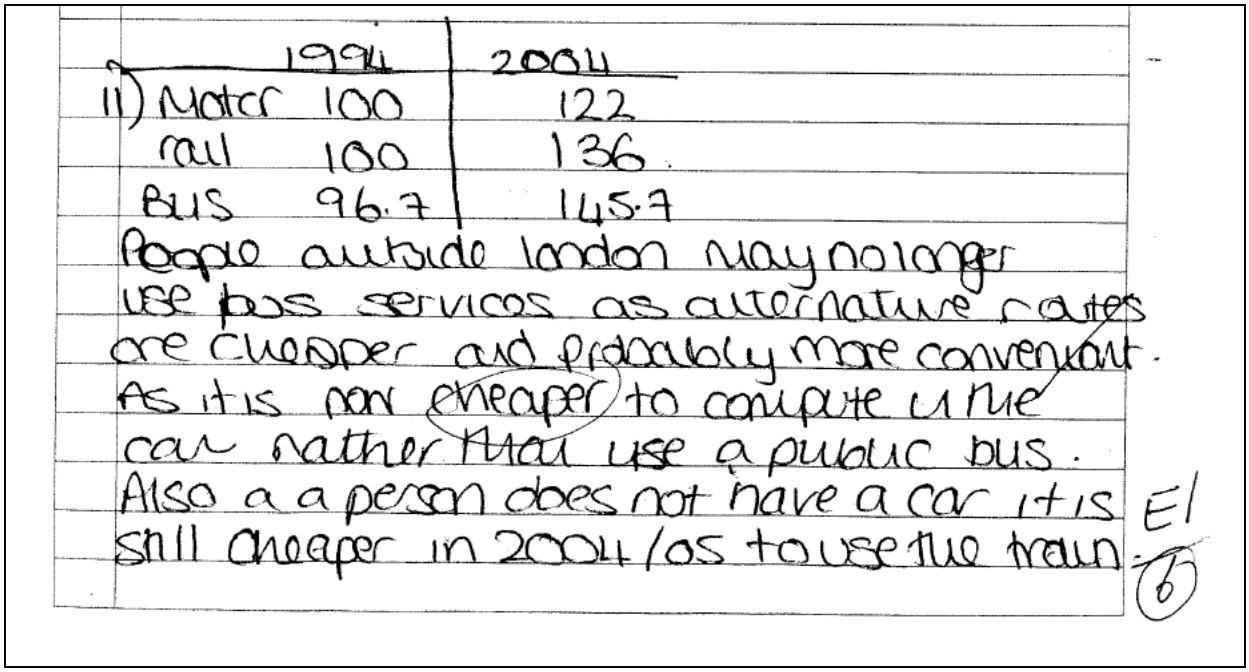
Q3 a 113 million ~~(113,000,000,000)~~ 2

b i) The trend over 1994/95 to 2004/05 is increasing year on year, as more passengers by the Million are using the local bus service

Leave blank

ii. The trend outside of london is year on year the numbers are falling. Showing less people use the service. 2
Therefore more people are using buses within London during 94 to 05 than people outside london during the same time period.

c i) A possible reason for the trend within london is that fares where slightly increasing hardly noticable for todays commuter. where as atside London fare where substantially increasing rising from 96.2 in 1994/95 to £152.5 in 2004/05. compared with london 96.2 (same) and 126.8 (Lower) E1



	1994	2004
ii) Motor	100	122
rail	100	136
BUS	96.7	145.7

People outside london may no longer use bus services as alternative routes are cheaper and probably more convenient. As it is now cheaper to compute in the car rather than use a public bus. Also a a person does not have a car it is still cheaper in 2004/05 to use the train. E1 6

Commentary

The candidate gained full marks for part (a).
In part (b) she correctly identified an upward trend in London and a downward trend outside London. One further point was required for full marks and her comment that more people are using buses in London than outside London is incorrect.
In part (c)(i) she correctly identifies that fares were rising faster outside London than inside London but seems to confuse the index with the actual fare. A comparison with the Retail Price Index was needed for full marks.
She picked up a mark in part (c)(ii) but again confused the index with the actual cost.

Mark Scheme

3(a)	113 million	B1 B1	 2	113 113 million
(b)(i)	Upward trend in London - relatively slow 1994/95 (decrease in 1998/99) - increasing more rapidly 2000/01 onwards.	E1		E1 upward in London
(ii)	Outside London there is a slow downward trend (apart from 1998/99 to 2000/01 when there was little change).	E1 E1	 3	E1 downward outside London E1 additional valid (but not trivial) point
(c)(i)	Increase in fares index outside London far exceeds increase in RPI - this explains reduction in bus journeys outside London.	E1		E1 comparison of fares outside London with RPI
	Increase in fares index in London is similar to increase in RPI. Thus any reason for increased bus journeys (eg congestion charge / increased population) should not be inhibited by price.	E1 E1	 3	E1 comparison of fares in London with RPI E1 comparison of London with outside London E1 any sensible conclusion **Maximum 3**
(ii)	Outside London increase in bus fares > increase in rail fares > both RPI and increase in motoring costs (which have declined in real terms).			
	This may explain reduction in bus journeys.	E1		E1 valid comparison outside London
	In London increase in bus fares is < increase in rail fares / similar to RPI / only slightly > than increase in motoring costs.			
	This may explain increase in bus journeys.	E1	2	E1 valid comparison in London
	Total		**10**	

Question 4

[A sheet of graph paper is provided for use in this question.]

A professional institution provides a statistics examination in January and June each year. The table shows the number of candidates who have taken the examination between June 2005 and January 2008.

	June 2005	**Jan 2006**	**June 2006**	**Jan 2007**	**June 2007**	**Jan 2008**
Number of candidates	1188	192	1351	238	1499	290

(a) Plot the data. **Your horizontal axis should extend to June 2009.** *(3 marks)*

(b) Calculate values of a suitable moving average and add them to your graph. *(4 marks)*

(c) Draw a trend line and estimate the seasonal effect for June. *(4 marks)*

(d) Estimate the number of candidates taking the examination in June 2009. Explain the method that you use. *(3 marks)*

(e) In June 2007, four examiners were used to mark the papers. It is estimated that, for any particular examination, an examiner can mark up to 400 papers.

Estimate how many examiners will be needed to mark the papers in June 2009. Justify your answer. *(2 marks)*

(f) Jackie fitted a regression equation to the numbers of candidates taking the examination in June 2005, June 2006 and June 2007. Using this equation, she forecast that about 1800 candidates would take the examination in June 2009.

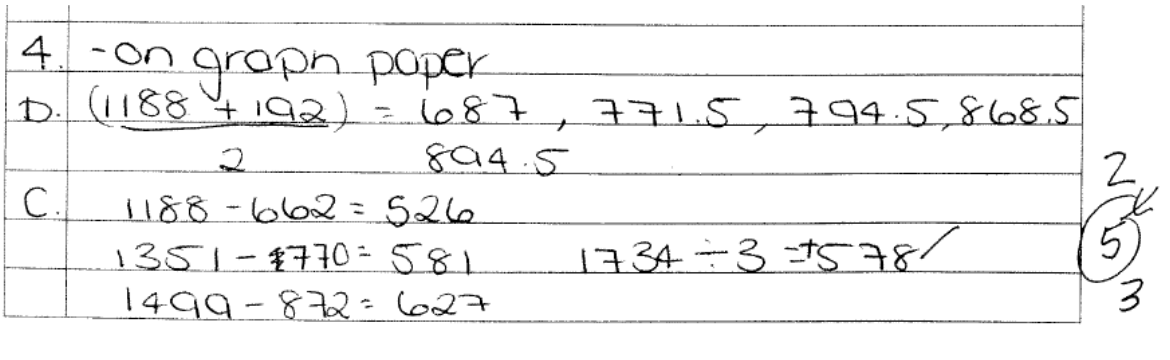
Comment on:

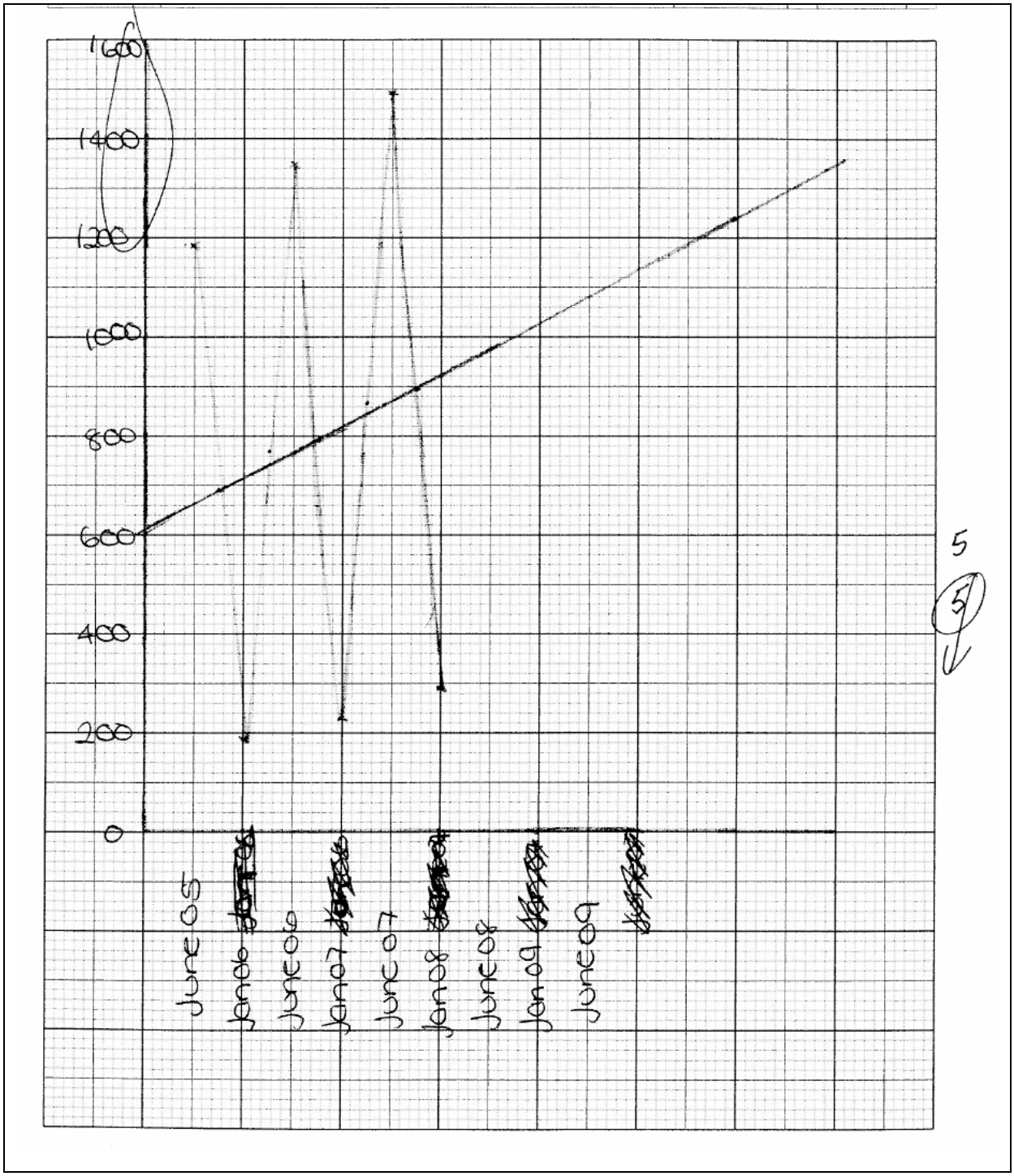
(i) whether this is a sensible method of forecasting;

(ii) the effect, if any, that this forecast would have on your answer to part (e).

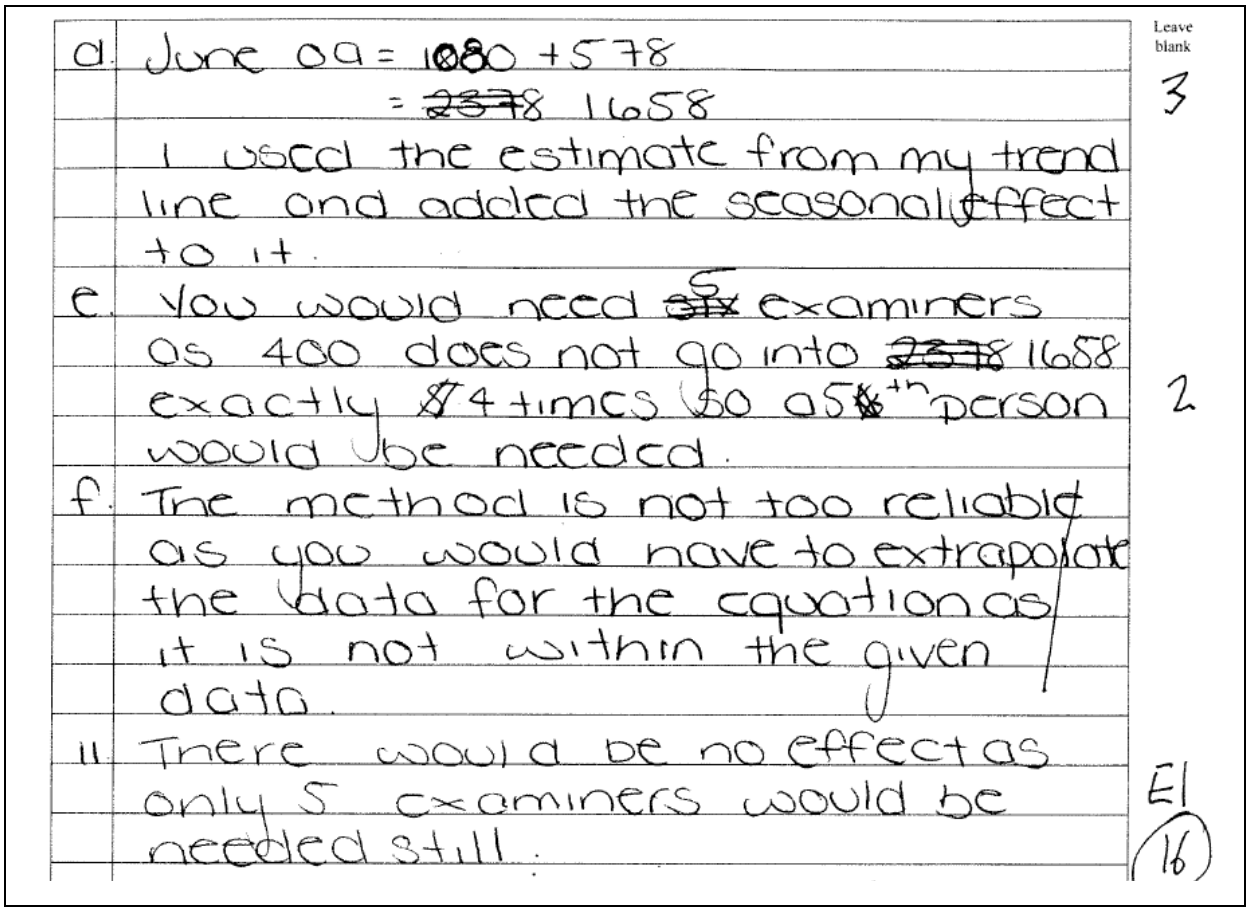
(3 marks)

Student Response




1600
1400
1200
1000
800
600
400
200
0
June 05
June 06
June 07
June 08
June 09

5
5



Commentary

A good answer. The candidate lost a mark for failing to label the vertical axis of her graph. It is good that in part (f)(i) the candidate demonstrates that she understands the dangers of extrapolation. However extrapolation is unavoidable when making a forecast and so cannot be given as a valid reason for criticising a particular method of forecasting.

Mark Scheme

4(a)	On graph	B1 M1 A1	 3	Scales labelled Method for plot Accurate plot by eye - allow one small slip. Disallow very small scale
(b)	Jun '05 Jan '06 Jun '06 Jan '07 Jun '07 Jan '08 1188 192 1351 238 1499 290 m.a. 690.0 771.5 794.5 868.5 894.5	B1 M1	 	2-point moving average used Method for moving average
	On graph	m1 A1	 4	Moving average plotted in correct position – must be 2-point Moving average plotted accurately – allow one small slip
(c)	Trend line on graph	B1		
	June effect (500+570+610)/3 = 560	M1 m1 A1	 4	Method for seasonal effect - ignore sign - may be earned in (d) Method for seasonal effect – allow '06 and '07 only 560 (530 ~ 600) - may be earned in (d)
(d)	Estimate of m.a. June 2009 = 1100	B1		1060 ~ 1140
	Estimated number of candidates = 1100 + 560 = 1660	M1 A1	 3	Method for forecast - their figures for positive seasonal effect 1660 (1620 ~ 1720) **SC** allow B1 for in range, no working
(e)	Max for 4 examiners is $4 \times 400 = 1600$	M1		Method - their figures
	Estimate of 1660 suggests 5 examiners will be required.	A1	2	CAO **SC** allow B1 for 5 without explanation
(f)(i)	Limited data available suggests January and June figures are both increasing approximately linearly but at different rates. Using regression on June figures is a sensible method but current trends may not continue.	E1 E1	 	Limited data; Jan and June increasing at different rates; sensible method; current trends may not continue **any 2 valid points**
(ii)	1800 would still need 5 examiners – no effect	E1√	3	ft no effect - their figures
			19	

Question 5

An ambulance service serves a large rural area. A sample of the times taken, in minutes, from receipt of an emergency call to the arrival of an ambulance at the location of the emergency was recorded as follows.

23 14 27 15 12 26 20 18 22

(a) Assume that these times may be regarded as a random sample from a normal distribution with standard deviation 3.5 minutes. Carry out a hypothesis test, at the 5% significance level, to examine whether the mean time could be equal to 19 minutes. *(8 marks)*

(b) It is now decided to examine, at the 5% significance level, whether the mean time is less than 19 minutes. State the:

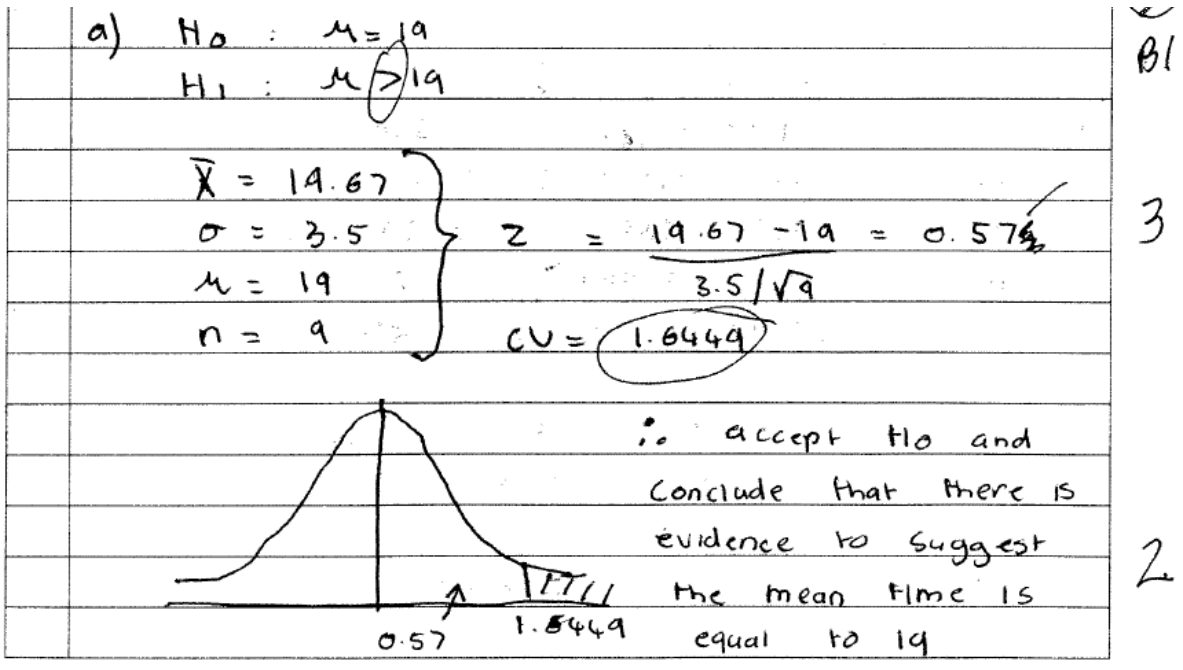
(i) hypotheses;

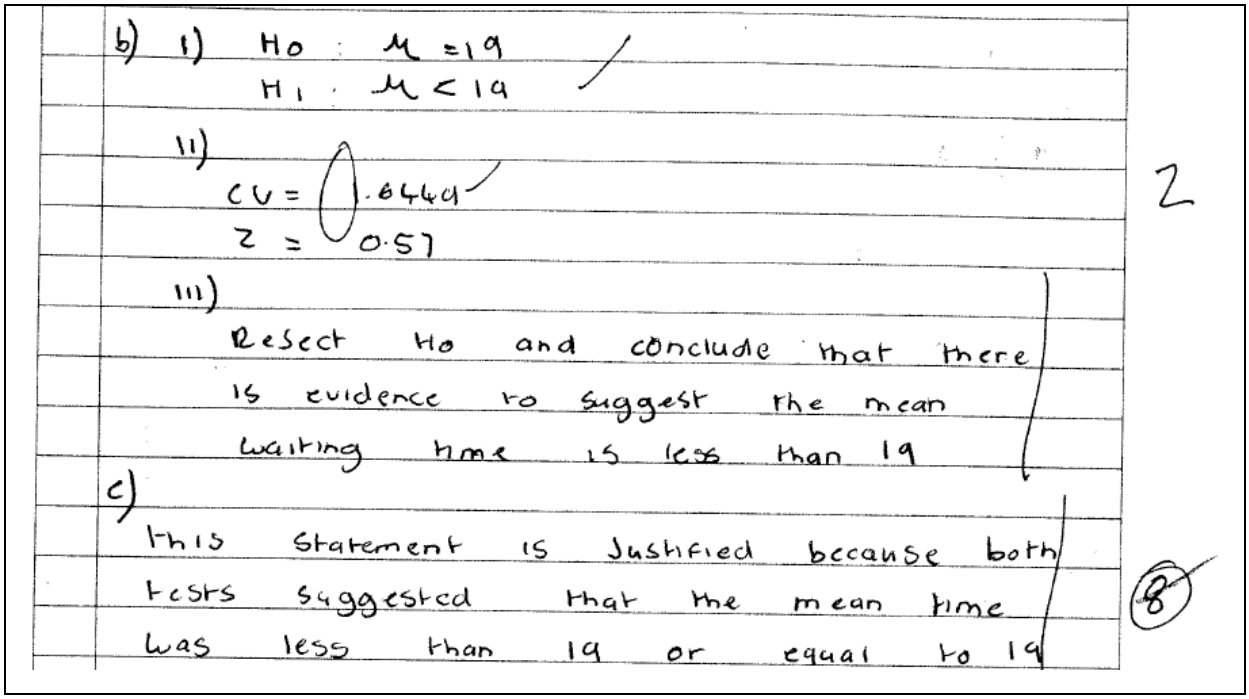
(ii) critical value(s);

(iii) conclusion. *(4 marks)*

(c) There is a target of at most 19 minutes for the mean time from receipt of an emergency call to the arrival of an ambulance at the location of the emergency. The director of the ambulance service states that the data provide significant evidence that this target has been achieved. Comment on this statement. *(3 marks)*

Student Response





b) i) $H_0: \mu = 19$
$H_1: \mu < 19$ ✓

ii) CV = −1.6449 ✓
z = 0.57

2

iii) Reject Ho and conclude that there is evidence to suggest the mean waiting time is less than 19

c) This statement is justified because both tests suggested that the mean time was less than 19 or equal to 19

8

Commentary

The candidate made a good attempt at part (a), but lost marks because he carried out a one-sided test instead of a two-sided test.
In part (b) the critical value should have been -1.6449 and no marks were given in part (b)(iii) unless the z-value was compared with a negative critical value.
The answer to part (c) is consistent with their (incorrect) conclusion in part (b) but as this incorrect conclusion makes the question easier to answer no marks were awarded.

Mark Scheme

5(a)	H_0: $\mu = 19$ H_1: $\mu \neq 19$	B1 B1		B1 one correct hypothesis B1 both hypotheses correct
	$\bar{x} = 19.667$ $z = (19.667 - 19)/(3.5/\sqrt{9}) = 0.571$	M1 m1 A1		Use of $3.5/\sqrt{9}$ Method for z – ignore sign 0.571 (0.57 ~ 0.575)
	c.v. ± 1.96 Accept H_0	B1 A1√		± 1.96 – ignore sign Conclusion – must be compared with correct tail of normal
	Conclude that there is no significant evidence that the mean time for ambulance to arrive is not 19 minutes	A1√	8	In context – needs previous A1√
(b)(i)	H_0: $\mu = 19$ H_1: $\mu < 19$	B1		Both hypotheses – ignore any errors already penalised in (a)
(ii)	–1.6449	B1 B1		1.6449 (1.64 ~ 1.65) Any negative z-value
(iii)	No significant evidence that the mean time for ambulance to arrive is less than 19 minutes.	A1√	4	Needs m mark in (a) and – c.v.
(c)	No significant evidence that target has been achieved.	E1		E1 director's comment incorrect
	Indeed as $\bar{x} = 19.66$ there is no evidence at all.	E1		E1 sample mean greater than 19
	There is however no significant evidence that it has not been achieved.	E1	3	E1 no significant evidence target has not been achieved. E1 no significant evidence target has been achieved **Maximum 3**
	Total		**15**	

Question 6

John organises the house-to-house delivery of free newspapers in a small town. He divides the town into areas and employs students to deliver the newspapers to the houses in each area. After a delivery, John calls on eight houses in each area to check that the free newspaper has been delivered.

A particular area consists of:

North Street, which contains 63 houses;
East Street, which contains 77 houses;
South Street, which contains 46 houses;
West Street, which contains 94 houses.

(a) Describe how John could select a simple random sample of eight houses from this area.
(5 marks)

(b) Describe how John could select a systematic sample of eight houses from this area.
(3 marks)

(c) John decides to select two of the four streets in this area at random and to call at four houses selected randomly from each of these two streets. Name this type of sampling.
(1 mark)

(d) Socrates is a student who is sometimes employed by John to deliver the free newspapers in this area. If Socrates does not have time to deliver to all the houses, he misses out one whole street.

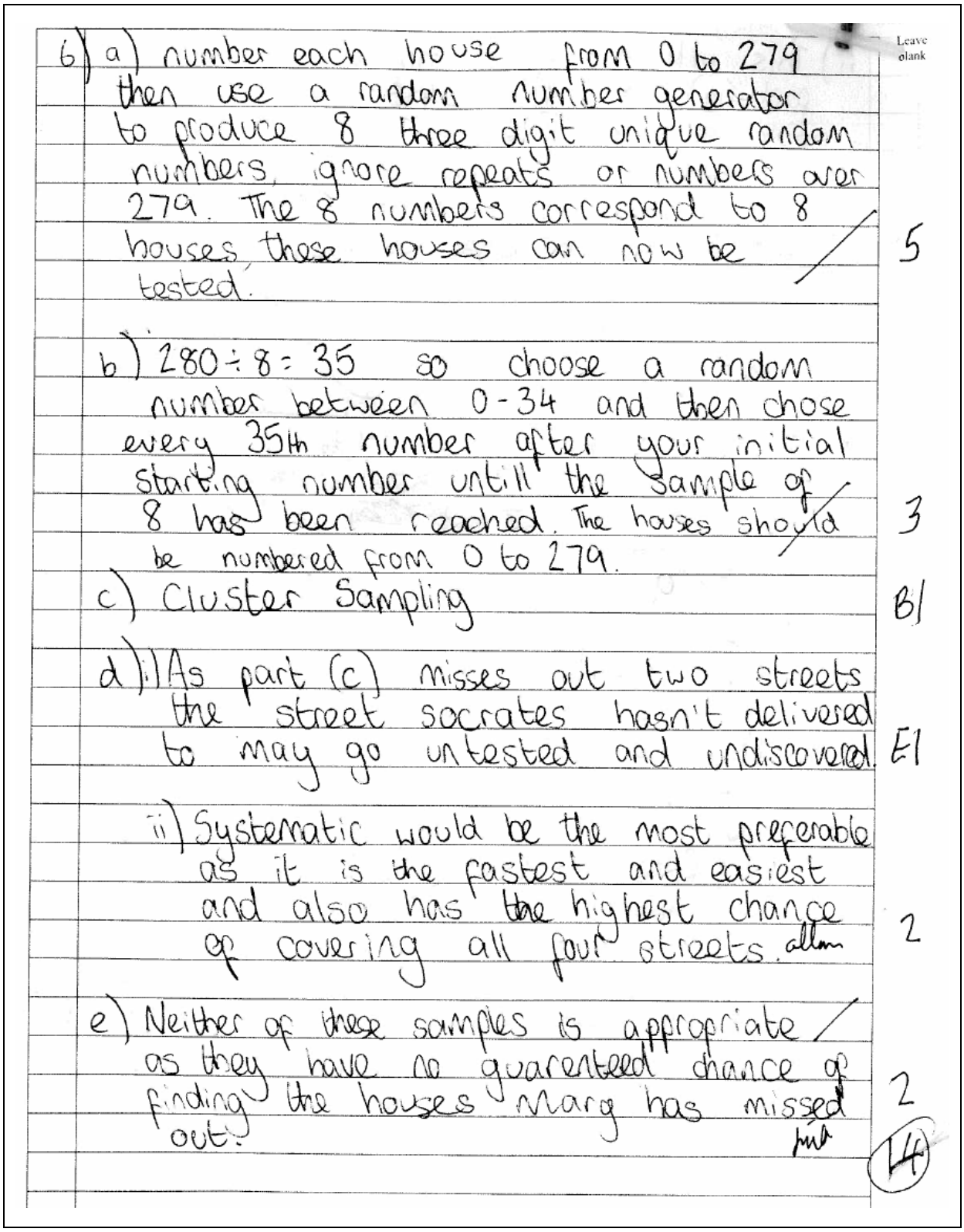
(i) Explain why John should not use the type of sampling described in part (c) when checking that Socrates has delivered to all houses in this area.

(ii) State, giving a reason, which, if either, of a simple random sample or a systematic sample would be preferable when checking that Socrates has delivered to all houses in this area.
(3 marks)

(e) Mary is another student who is sometimes employed by John to deliver the free newspapers in this area. If Mary does not have time to deliver to all the houses, she misses out houses haphazardly.

State, giving a reason, which, if either, of a simple random sample or a systematic sample would be preferable when checking that Mary has delivered to all houses in this area.
(2 marks)

Student Response



6) a) number each house from 0 to 279 then use a random number generator to produce 8 three digit unique random numbers, ignore repeats or numbers over 279. The 8 numbers correspond to 8 houses these houses can now be tested. 5

b) 280 ÷ 8 = 35 so choose a random number between 0-34 and then chose every 35th number after your initial starting number until the sample of 8 has been reached. The houses should be numbered from 0 to 279. 3

c) Cluster Sampling B1

d) i) As part (c) misses out two streets the street socrates hasn't delivered to may go untested and undiscovered. E1

ii) Systematic would be the most preferable as it is the fastest and easiest and also has the highest chance of covering all four streets. allow 2

e) Neither of these samples is appropriate as they have no guarenteed chance of finding the houses Marg has missed out. 2

14

Commentary

A very good answer. It is unusual for a candidate to score full marks on the sampling question and this candidate has done all that could be asked of them.

Mark Scheme

6(a)	280 houses	B1		
	Number houses 000 to 279	E1		OE - their total
	Select 3-digit random numbers	E1		
	Ignore repeats	E1		
	and > 279	E1	5	Consistent with their numbering
	Continue until 8 numbers obtained Select corresponding houses			
(b)	Number houses street by street, eg North St 000–062 East St 063–139 South St 140–185 West St 186–279	E1		E1 number houses street by street - may be earned in (a) but more detail required here E1 idea of systematic sampling
	Select a random number between 00 and 34.	E1		E1 choose random starting point
	Choose this house and every 35th house thereafter.	B1	3	B1 every 35th house **Maximum 3**
(c)	Cluster	B1	1	
(d)(i)	If Socrates misses a street there is a substantial probability (0.5) that John will not check any houses in this street.	E1		
(ii)	Systematic preferred	B1		
	John certain to check some houses in each street	E1	3	
(e)	No preference	B1		
	Both equally likely to check houses missed by Mary	E1	2	**SC** allow B1 for systematic because easier to carry out
	Total		**14**	