Pearson Edexcel

Mark Scheme (Topic Test)

Sampling and Data Collection

Pearson Edexcel GCE
In Statistics (9ST0)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

General Marking Guidance

Total marks

The total number of marks for the paper is 80.

Mark types

The Edexcel Statistics mark schemes use the following types of marks:

- **M** **Method** marks,
  awarded for 'knowing a method and attempting to apply it',
  unless otherwise indicated.
- **A** **Accuracy** marks can only be awarded if the relevant method (M) marks have been earned.
- **B** **Unconditional accuracy** marks are independent of M marks
- **E** **Explanation** marks

NOTE: Marks should not be subdivided.

Abbreviations

These are some of the marking abbreviations that will appear in the mark schemes.

- ft follow through
- PI possibly implied
- cao correct answer only
- cso correct solution only
  (There must be no errors in this part of the question)
- awrt answers which round to
- awfw answers which fall within (a given range)
- SC special case
- nms no method shown
- oe or equivalent
- dep dependent (on a given mark or objective)
- dp decimal places
- sf significant figures
- ✱ The answer is printed on the paper

Further notes

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied **positively**. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is **no ceiling** on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- All A marks are 'correct answer only' (cao), unless shown, for example, as A1ft to indicate that previous wrong working is to be followed through.
- All M marks are 'possibly implied' (PI) unless specifically stated otherwise in the 'Notes' column.
- After a **misread**, the subsequent A marks affected are treated as A1ft, but manifestly absurd answers should never be awarded A marks.
- **Crossed out** work should be marked UNLESS the candidate has replaced it with an alternative response.
- If **two solutions** are given, each should be marked, and the resultant mark should be the mean of the two marks, rounded down to the nearest integer if needed.

Question	Scheme	Marks	AO	Notes
1(a)	Snowball (sampling)	B1	1.1	
1(b)	**Possible advantages (not exhaustive)**			
	Easy / quick to set up			
	Social media is used by a lot of people			
	More likely to receive relevant responses since people who have not experienced the restaurant are less likely to answer			
	Allowing people to share the link allows the survey to reach people who may not use social media			
		E1	3.1a	Any reasonable advantage
1(c)	**Possible disadvantages (not exhaustive)**			
	Since it is online, only people with access to the internet can respond			
	People who have never been to the restaurant may fill in the survey giving a false response			
	Relies on participants to respond			
	Relies on participants to respond truthfully			
	Snowball sampling is not random / any data collected may not be suitable for statistical testing.			
		E1	3.1a	Any reasonable disadvantage
	Total	**3**		

Question	Scheme	Marks	AO	Notes
2(a)	Judgemental (sampling)	B1	1.1	
2(b)	**Possible advantages (not exhaustive)**			
	Easy / quick to collect data			
	The shopping centre is large and there is no need for a sampling frame / list of people who attend the shopping centre			
	The shopping centre is likely to contain people who have an opinion on the town's public toilet facilities			
		E1	3.1a	Any reasonable advantage
	Possible disadvantages (not exhaustive)			
	Biased towards people who go to the shopping centre on that day			
	The shopping centre may contain people from out of town who do not use the public toilet facilities in the town			
	The people in the shopping centre may not be representative of the town's population			
		E1	3.1a	Any reasonable disadvantage
	Total	**3**		

Question	Scheme	Marks	AO	Notes
3	The **running times** of the 100 m sprint should be collected	E1	1.1	
	Possible sampling methods (not exhaustive)			
	Take a (simple) random sample of athletes in the country and group them by running club			
	Take a stratified sample of athletes stratified by running club			Proportional or disproportional
	Take a cluster sample of (e.g.) 5 running clubs			
	Take a judgemental sample of 100 metre sprinters from each running club			
		E1	1.1	Any reasonable sampling method which results in at least two groups of athletes from different running clubs Must be in context but isw about specific details of collecting the data
	Total	**2**		

Question	Scheme	Marks	AO	Notes
4	Generate random numbers in the range 1 to 25	E1	1.1	Or 2-digit numbers, ignoring 00 and above 25
	Do not ignore repeats	E1	1.1	Must be stated
	Select the audience members corresponding to these numbers	E1	1.1	Must be in context
	Total	**3**		

Question	Scheme	Marks	AO	Notes
5(a)	Number the students from 1 to 1200	E1	1.1	oe any set of 1200 consecutive numbers
	Generate a random number in the range 1 to 20	E1	1.1	Or first 20 numbers in range given in first E1
	Select the student corresponding to this number and the students corresponding to every 20^{th} number after that	E1	1.1	Must be in context
5(b)	Yes…	B1	1.1	Independent of reason
	…because the sample size divides the population size	E1	2.1b	oe e.g. 1200/60 is an integer
5(c)	No…	B1	1.1	Independent of reason
	…because not every sample of 60 students can be selected	E1	2.1b	oe e.g. "a sample containing students with numbers 1 to 60 can be chosen" Must be in context
	Total	**7**		

Question	Scheme	Marks	AO	Notes
6(a)	Generate random numbers in the range 1 to 639	E1	1.1	Or three digit numbers ignoring 000 and above 639
	Ignore repeats	E1	1.1	
	Select the (molar) teeth corresponding to these numbers	E1	1.1	Must be in context
6(b)	There is no bias in the sampling method	B1	1.1	Accept "no"
6(c)	The data collected are biased towards molar teeth / extracted teeth and may not be representative of all teeth	E1	2.1b	Must be in context
	Total	**5**		

Question	Scheme	Marks	AO	Notes
7	$\frac{60}{60+660} \times 72 = 6$ smokers/vapers	M1	1.2	Attempt to find proportion of population for either group
	$\frac{660}{60+660} \times 72 = 66$ non smokers/vapers	A1	1.2	For 6 smokers/vapers
		A1	1.2	For 66 non-smokers/vapers
	Assign the smokers/vapers a number from 1 to 60	E1	1.1	Or 1 to 660 for non-smokers Oe any range of 60 numbers
	Use a random number generator to generate numbers in the range 1 to 60, (ignoring repeats)	E1	1.1	Or range stated above
	Select the smokers/vapers corresponding to these numbers	E1	1.1	Or "non-smokers/vapers"
	Assign the non-smokers/vapers a number from 1 to 660 Use a random number generator to generator numbers in the range 1 to 660, ignoring repeats. Select the non-smokers/vapers corresponding to these numbers	E1	1.1	Indication of the repeat process for the stratum not described for first three E1 Does not have to be explained in full, but number ranges must be stated e.g. "repeat the process for the non-smokers/vapers using the range 1 – 660"
	Total	**7**		

Question	Scheme	Marks	AO	Notes
8(a)	**Suggestion A:** Cluster (sampling)	B1	1.1	
	Suggestion B: Simple random (sampling)	B1	1.1	
				Must be clearly identified as A and B
8(b)	**Suggestion A:** No	B1	1.1	Condone "Maybe" or "Yes" with a reasonable scenario e.g. "if the fast-food branches are selected at random and each branch employs the sample number of people"
	Suggestion B: Yes	B1	1.1	
				Must be clearly identified as A and B
8(c)	**Possible reasons (not exhaustive)**			
	More convenient if the restaurants are chosen for convenience			
	The employees selected from the five restaurants are equally weighted			
	Only need a sampling frame / list of employees of the 5 restaurants chosen			
		E1	3.1a	Must be an advantage of Cluster over Simple Random in context
	Possible reasons (not exhaustive)			
	The sample of employees will be random			
	More likely to obtain a sample representative of all employees			
		E1	3.1a	Must be an advantage of Simple Random over Cluster in context
	Total	**6**		

Question	Scheme	Marks	AO	Notes
9(a)(i)	**Suggestion A:** Cluster (sampling)	B1	1.1	
	Suggestion B: Random (Sampling)	B1	1.1	May be simple or unrestricted
	Suggestion C: Stratified (Sampling)	B1	1.1	Condone "Disproportional" for this mark
	Suggestion D: Systematic (Sampling)	B1	1.1	
				Must be clearly identified to the correct suggestion
9(a)(ii)	**Suggestion A:** No	B1	1.1	Condone "Yes" or "Maybe" if a reasonable scenario if described e.g. "if the 4 neighbourhoods or villiages are chosen at random and there is an equal number in each neighbourhood / village"
	Suggestion B: Yes	B1	1.1	
	Suggestion C: Yes	B1	1.1	
	Suggestion D: Yes	B1	1.1	
				Must be clearly identified to the correct suggestion
9(b)	No	B1	1.1	
	It is not possible to obtain (for example) a sample of electors all from urban areas	E1	2.1b	Oe must be in context
9(c)	Yes	B1	1.1	May be seen in (a)(i)
	The sample of electors will be fairly represented across urban, suburban and rural areas	E1	3.1a	Oe indication of equal representation
9(d)	Suggestion C will definitely have electors who live in urban areas but most of the sample will definitely have electors who do not live in urban areas. May not be useful.	E1	2.1b	Must have a reference to usefulness with a valid reason
	Suggestion B may have lots of electors who live in urban areas but also may not have any electors who live in urban areas. May not be useful.	E1	2.1b	Must have a reference to usefulness with a valid reason
				Ignore any comments relating to which method is more useful
	Total	**14**		