Pearson Edexcel

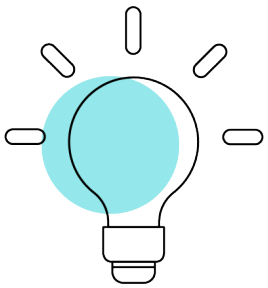
Topic Test
Summer 2022

Pearson Edexcel GCE Mathematics (9MA0)

Paper 3 – Statistics

Topic 2: Measures of central tendency and variation

Contents



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 guidance to Topic Tests

Context

- Topic Tests have come from past papers both published (those materials available publicly) and unpublished (those currently under padlock to our centres) presented in a different format to allow teachers to adapt them for use with candidates.

Purpose

- The purpose of this resource is to provide qualification-specific sets/groups of questions covering the knowledge, skills and understanding relevant to this Pearson qualification.
- This document should be used in conjunction with the advance information for the subject as well as general marking guidance for the qualification (available in published mark schemes).

Revise Revision Guide content coverage

The questions in this topic test have been taken from past papers, and have been selected as they cover the topic(s) most closely aligned to the A level advance information for summer 2022:

- Topic 2: Measures of central tendency and variation

The focus of content in this topic test can be found in the Revise Pearson Edexcel A level Mathematics Revision Guide. Free access to this Revise Guide is available for front of class use, to support your students' revision.

Contents	Revise Guide page reference	Level
Pure Mathematics	1-111	A level
Statistics	112-147	A level
Mechanics	148-181	A level

Content on other pages may also be useful, including for synoptic questions which bring together learning from across the specification.

Questions

Question T2_Q1

4. Charlie is studying the time it takes members of his company to travel to the office. He stands by the door to the office from 08 40 to 08 50 one morning and asks workers, as they arrive, how long their journey was.

(a) State the sampling method Charlie used.

(1)

(b) State and briefly describe an alternative method of non-random sampling Charlie could have used to obtain a sample of 40 workers.

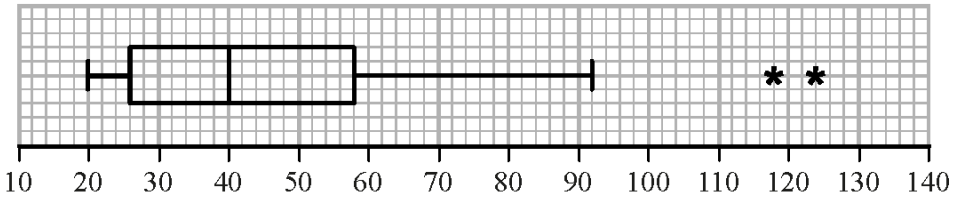
(2)

Taruni decided to ask every member of the company the time, x minutes, it takes them to travel to the office.

(c) State the data selection process Taruni used.

(1)

Taruni's results are summarised by the box plot and summary statistics below.



Journey time (minutes)

$n = 95 \quad \sum x = 4133 \quad \sum x^2 = 202\,294$

(d) Write down the interquartile range for these data.

(1)

(e) Calculate the mean and the standard deviation for these data.

(3)

(f) State, giving a reason, whether you would recommend using the mean and standard deviation or the median and interquartile range to describe these data.

(2)

Rana and David both work for the company and have both moved house since Taruni collected her data.

Rana's journey to work has changed from 75 minutes to 35 minutes and David's journey to work has changed from 60 minutes to 33 minutes.

Taruni drew her box plot again and only had to change two values.

(g) Explain which two values Taruni must have changed and whether each of these values has increased or decreased.

(3)

Question 4 continued

Question 4 continued

Question 4 continued

Question T2_Q2

2.

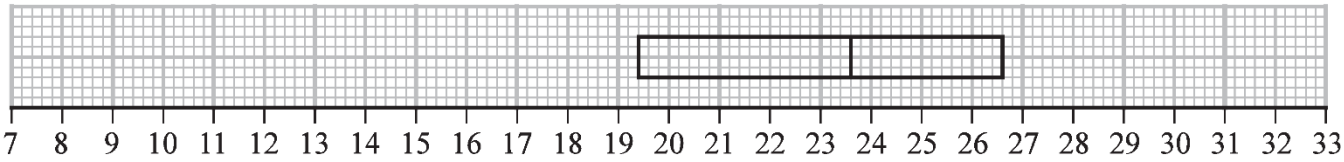


Figure 1

The partially completed box plot in Figure 1 shows the distribution of daily mean air temperatures using the data from the large data set for Beijing in 2015

An outlier is defined as a value
more than 1.5 × IQR below Q_1 or
more than 1.5 × IQR above Q_3

The three lowest air temperatures in the data set are 7.6 °C, 8.1 °C and 9.1 °C
The highest air temperature in the data set is 32.5 °C

(a) Complete the box plot in Figure 1 showing clearly any outliers.

(4)

(b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come.

(1)

Using the data from the large data set, Simon produced the following summary statistics for the daily mean air temperature, x °C, for Beijing in 2015

$$n = 184 \qquad \sum x = 4153.6 \qquad S_{xx} = 4952.906$$

(c) Show that, to 3 significant figures, the standard deviation is 5.19 °C

(1)

Simon decides to model the air temperatures with the random variable

$$T \sim N(22.6, 5.19^2)$$

(d) Using Simon's model, calculate the 10th to 90th interpercentile range.

(3)

Simon wants to model another variable from the large data set for Beijing using a normal distribution.

(e) State two variables from the large data set for Beijing that are **not** suitable to be modelled by a normal distribution. Give a reason for each answer.

(2)

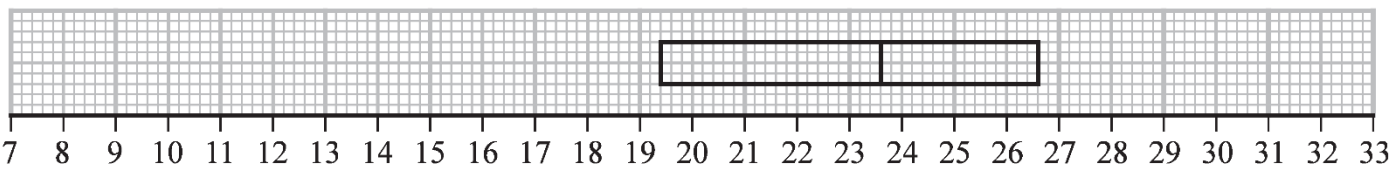
Question 2 continued

Question 2 continued

Turn over for a spare grid if you need to redraw your box plot.

Question 2 continued

Only use this grid if you need to redraw your box plot.

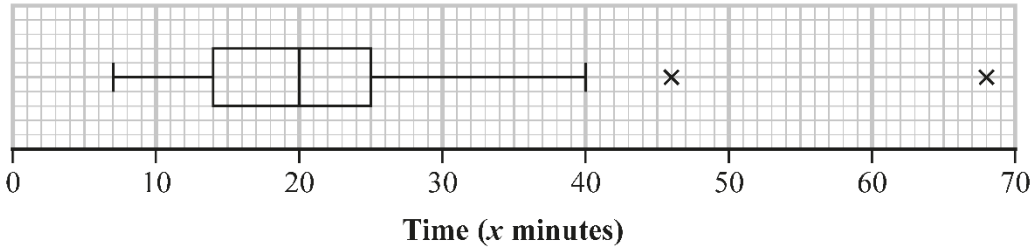


Question T2_Q3

3. Each member of a group of 27 people was timed when completing a puzzle.

The time taken, x minutes, for each member of the group was recorded.

These times are summarised in the following box and whisker plot.



Time (x minutes)

(a) Find the range of the times.

(1)

(b) Find the interquartile range of the times.

(1)

For these 27 people $\sum x = 607.5$ and $\sum x^2 = 17\,623.25$

(c) calculate the mean time taken to complete the puzzle,

(1)

(d) calculate the standard deviation of the times taken to complete the puzzle.

(2)

Taruni defines an outlier as a value more than 3 standard deviations above the mean.

(e) State how many outliers Taruni would say there are in these data, giving a reason for your answer.

(1)

Adam and Beth also completed the puzzle in a minutes and b minutes respectively, where $a > b$.
When their times are included with the data of the other 27 people

- the median time increases
- the mean time does not change

(f) Suggest a possible value for a and a possible value for b, explaining how your values satisfy the above conditions.

(3)

(g) Without carrying out any further calculations, explain why the standard deviation of all 29 times will be lower than your answer to part (d).

(1)

Question 3 continued

Question 3 continued

Question 3 continued

Question T2_Q4

3. Stav is studying the large data set for September 2015

He codes the variable Daily Mean Pressure, x, using the formula $y = x - 1010$

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \qquad \sum y^2 = 5912$$

(a) State the units of the variable x

(1)

(b) Find the mean Daily Mean Pressure for these 30 days.

(2)

(c) Find the standard deviation of Daily Mean Pressure for these 30 days.

(3)

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure
- in an **anticlockwise** direction around a region of **low** pressure

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015

Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1029	1028	1028
Cardinal Wind Direction			

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

W NE E

You may assume that these 3 locations were under a single region of pressure.

(d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table.
Give a reason for your answer.

(2)

Question 3 continued.

Mark Scheme

Question T2_Q1

Qu 4	Scheme	Marks	AO
(a)	Convenience or opportunity [sampling]	B1 (1)	1.2
(b)	Quota [sampling]	B1	1.1a
	e.g. Take 4 people every 10 minutes	B1 (2)	1.1b
(c)	Census	B1 (1)	1.2
(d)	[58 – 26 =] **32** (min)	B1 (1)	1.1b
(e)	$\mu = \frac{4133}{95} = 43.505263\ldots$ **awrt 43.5 (min)**	B1	1.1b
	$\sigma_x = \sqrt{\frac{202\,294}{95} - \mu^2} = \sqrt{236.7026\ldots}$	M1	1.1b
	= 15.385… **awrt 15.4 (min)**	A1 (3)	1.1b
(f)	There are outliers in the data (or data is skew) which will affect mean and sd	B1	2.4
	Therefore use median and IQR	dB1 (2)	2.4
(g)	Value of 20, LQ at 26 and outliers will not change or state that median and upper quartile are the values that do change	B1	1.1b
	More values now below 40 than above so Q_2 or Q_3 will change and be lower	M1	2.1
	Both Q_2 and Q_3 will be lower	A1 (3)	2.4
		(13 marks)	

	Notes
(b)	1st B1 for quota (sampling) mentioned ("Stratified" or "systematic" or "random" are B0B0) 2nd B1 for a description of how such a system might work, requires suitable strata or categories e.g. time slots, departments, gender, age groups, distance travelled etc Suggestion of randomness is B0
(e)	B1 for a correct mean (awrt 43.5) M1 for a correct expression for the sd (including $\sqrt{}$)ft their mean A1 for awrt 15.4 (Allow s = 15.4667… awrt 15.5)
(f)	1st B1 for acknowledging outliers or skewness are a problem for mean and sd "extreme values"/"anomalies" OK May be implied by saying median and IQR not affected by.. We need to see mention of "outliers", "skewness" and the problem so "data is skewed so use median and IQR" is B0 unless mention that they are not affected by extreme values or mean and standard deviation can be "inflated" by the positive skew etc 2nd dB1 dep on 1st B1 for therefore choosing median and IQR
(g)	B1 for identifying 2 of these 3 groups of unchanged values or stating only Q_2 and Q_3 change M1 for explaining that median or UQ should be lower. E.g. the 2 values have moved to below 40 (or 58) and therefore more than 50% below 40 or (more than 75% below 58) or an argument to show that the other 3 values are the same. (o.e.) Allow arrows on box plot provided statement in words about increased % below 40 or 58 etc A1 for stating median and UQ are both lower with clear evidence of M1 scored [If lots of values on 40 then median might not change but, since two values do change then UQ would change. If this meant that 92 became an outlier then we would have a new value for upper whisker and an extra outlier so effectively 3 values are altered. So median changes]

Question T2_Q2

Question	Scheme		Marks	AOs
2(a)	IQR = 26.6 – 19.4 [= 7.2]		B1	2.1
	19.4 – 1.5 × '7.2' [= 8.6] or 26.6 + 1.5 × '7.2' [= 37.4]		M1	1.1b
	Plotting one upper whisker to 32.5 **and** one lower whisker to 8.6 or 9.1		A1	1.1b
	Plotting 7.6 and 8.1 as the only two outliers		A1	1.1b
			(4)	
(b)	October (since it is the month with the coldest temperatures between May and October in Beijing)		B1	2.4
			(1)	
(c)	$[\sigma =]\sqrt{\frac{4952.906}{184}}$ or e.g. $[\sigma =]\sqrt{\frac{S_{xx}}{n}} = 5.188...$ [=5.19*]		B1cso*	1.1b
			(1)	
(d)	$z = (\pm)\ 1.28(16)$	$[P_{90} =]29.251...$ or $[P_{10} =]15.948...$	B1	3.1b
	2 × 1.2816 × 5.19	'29.251…' – '15.948…'	M1	1.1b
		= awrt **13.3**	A1	1.1b
			(3)	
(e)	Daily mean wind speed/Beaufort conversion since it is qualitative Rainfall since it is not symmetric/lots of days with 0 rainfall		B1 B1	2.4 2.4
			(2)	
				(11 marks)

Notes	
(a)	**B1:** for a correct calculation for the IQR (implied by 10.8 **or** 8.6 **or** 37.4 seen) **M1:** for a complete method for either lower outlier limit or upper outlier limit (allow ft on their IQR) (may be implied by the 1st A1 or a lower whisker at 8.6) **A1:** both whiskers plotted correctly (allow ½ square tolerance) **A1:** only two outliers plotted, 7.6 and 8.1 (must be disconnected from whisker) **NOTE:** A fully correct box plot with no incorrect working scores 4/4
(c)	**B1cso*:** Correct expression with square root **or** correct formula and 5.188 or better Allow a complete correct method finding $\sum x^2 =$ awrt 98720 and $\sigma = \sqrt{\frac{98715.9...}{184} - \left(\frac{4153.6}{184}\right)^2}$
(d)	**B1:** Identifying z-value for 10th or 90th percentile (allow awrt (±) 1.28) or for identifying $[P_{90} =]29.251...$ (awrt 29.3) or $[P_{10} =]15.948...$ (awrt 15.9) (This may be implied by a correct answer awrt 13.3) **M1:** for $2 \times z \times 5.19$ where $1 < z < 2$ or for their $P_{90} - P_{10}$ where $25 < P_{90} < 35$ and $10 < P_{10} < 20$ **A1:** awrt 13.3
(e)	**B1:** for one variable identified and a correct supporting reason **B1:** for two variables identified and a correct supporting reason for each Allow any two of the following: • Wind speed/Beaufort since the data is non-numeric (o.e.). They need not mention Beaufort provided there is a description of the data as non-numeric (Do not allow wind direction/wind gust) • Rainfall as not symmetric/is skewed/is not bell shaped/lots of 0s /many days with no rain/mean≠mode or median • Date since each data value appears once/it is uniformly distributed • Daily mean pressure since it is not symmetric/is skewed/not bell shaped • Daily mean wind speed since it is not symmetric/is skewed/not bell shaped Do not allow 'not continuous' or 'discrete' as a supporting reason. Ignore extraneous non-contradicting statements

Question T2_Q3

Qu 3	Scheme	Marks	AO
(a)	[68 – 7 =] **61** (only)	B1 **(1)**	1.1b
(b)	[25 – 14] = **11**	B1 **(1)**	1.1b
(c)	$\left[\mu \text{ or } \bar{x} = \frac{607.5}{27} = \right]$ = **22.5**	B1 **(1)**	1.1b
(d)	$\sigma = \sqrt{\frac{17\,623.25}{27} - "22.5"^2}$ or $\sqrt{146.4629...}$	M1	1.1b
	= 12.10218… awrt **12.1**	A1 **(2)**	1.1b
(e)	$\mu + 3\sigma = "22.5" + 3 \times "12.1..."$ = awrt 59 so only **one** outlier	B1ft **(1)**	1.1b
(f)	Median increases implies that both values must be > 20	M1	3.1b
	Mean is the same means that $a + b = 45$	M1	1.1b
	So possible values are: e.g. $b = 21$ and $a = 24$ (o.e.)	A1 **(3)**	2.2b
(g)	Both values will be less than 1 standard deviation from the mean and so the standard deviation of all 29 values will be smaller	B1 **(1)**	2.4
		(10 marks)	

	Notes
(a)	B1 for correctly interpreting the box plot to find the range (more than 1 answer is B0)
(b)	B1 for correct understanding of IQR and answer of 11
(c)	B1 for 22.5 only (or exact equivalent such as $\frac{45}{2}$). Allow 22 mins and 30 secs.
(d)	M1 for a correct expression including square root. Allow $\sqrt{146}$ or better. Ft their mean A1 for awrt 12.1 NB Allow use of s = 12.3327… or awrt 12.3
(e)	B1ft for a correct calculation or value based on their μ and σ and compatible conclusion
(f)	1st M1 Correct start to the problem and a correct statement about the values based on median Allow if their final two values are both >20 2nd M1 for a correct explanation leading to equation $a + b = 45$ (o.e. e.g. equidistant from mean) Allow if their final two values sum to 45 A1 for a correct pair of values (both > 20 with a sum of 45) **and** at least some attempt to explain how their values satisfy at least one of the conditions (both > 20 or $a + b = 45$). Ignore $a =$ or $b =$ labels **NB** The values for a and b do not need to be integers.
(g)	B1 for a correct explanation. Must mention that both values are less than 1 sd (ft their answer to (d)) from the mean

Question T2_Q4

Qu 3	Scheme	Marks	AO
(a)	Hectopascal or hPa	B1 **(1)**	1.2
(b)	$\overline{x} = \overline{y} + 1010$ or $\frac{214}{30} + 1010$	M1	1.1b
	= 1017.1333… awrt **1017**	A1 **(2)**	1.1b
(c)	$\sigma_x = \sigma_y$ (or statement that standard deviation is not affected by this type of coding)	M1	3.1b
	$\left[\sigma_y =\right] \sqrt{\frac{5912}{30} - (\text{"7.13[33...]"})^2}$ or $\sqrt{146.1822...}$	M1	1.1b
	= 12.0905… awrt **12.1**	A1 **(3)**	1.1b
(d)	High pressure (since approx. mean + sd) so clockwise Locations are (from North to South): Leuchars, Heathrow, Hurn	B1	2.4
	Wind direction is direction wind blows from So: Heathrow (**NE**) Hurn (**E**) Leuchars (**W**)	B1 **(2)**	2.2a
		(8 marks)	

	Notes
FYI	1 hPa = 100 Pa; 10hPa = 1 kPa; 1Pa = 1 Nm^{-2}
(a)	B1 for "hectopascal" or hPa (condone pascals, allow millibars or mb) o.e. Do NOT allow kPa or kilopascals or Pa on its own
(b)	M1 for a strategy to find $\overline{x}$ Allow an attempt to find $\sum x$ that gets as far as $\sum x = \sum y + 30 \times 1010$ [= 30 514] A1 for awrt 1017 (accept 1020) [Ignore incorrect units]
(c)	1st M1 for an overall strategy using the fact $\sigma_x = \sigma_y$ (can be implied by correct final ans) or for $\sum x = 30\,514$ and $\sum x^2 = 31\,041\,192$ (both seen and correct) 2nd M1 for a correct expression (with $\sqrt{}$)(ft their $\overline{y}$ to 3sf) allow awrt 146 for 146.1822.. or for correct expression in x can ft their $\sum x > 30\,000$ or their answer to (b) A1 (dep on 2nd M1) for awrt 12.1 [Ignore incorrect units]
Final answer	Final ans of awrt 12.1 scores 3/3 **but** if they then adjust for x e.g. add 1010 (M0M1A1)
(d)	1st B1 for at least one of these reasons (these 2 lines) clearly stated (may see diagram) Need "high pressure" **and** "clockwise" to score on 1st line Contradictory statements B0 e.g. correct N~S list but say "anticlockwise" 2nd B1 (indep of 1st B1) for deducing the 3 correct directions either in the table or stated as above If the answers in table and text are different we take the table (as question says)