Grouped Data, Percentiles & Modal Interval (Grade 10)

MARKING GUIDELINE

MARKS	DURATION	QUESTIONS
26	40 minutes	2

Name: ..... Class: .....

Date: .....

MARK ____ / 26

INSTRUCTIONS AND INFORMATION

1. Answer **ALL** the questions in this question paper.
2. Answer QUESTION 1 by **circling** the letter (A–D) in the answer grid at the end of that section.
3. Show **ALL** calculations clearly.
4. Show all units where applicable.
5. Number the answers correctly according to the numbering system used in this question paper.
6. A non-programmable calculator may be used, unless stated otherwise.
7. Write neatly and legibly.

QUESTION 1 [10 MARKS]

Four options are given as possible answers to the following questions. Choose the correct answer and circle the letter (A–D) in the grid at the end of this section. If you want to change your choice, put a cross through the wrong letter and circle your new choice.

1.1 For the interval $20 \leq x < 30$, the MIDPOINT used to estimate the mean is ... (1)

A) 30 **B)** 20

C) 25 ✓ **D)** 10

> **Answer: C** — The midpoint is the average of the two endpoints: $(20 + 30) \div 2 = 25$.
>
> - **A** — that is the upper boundary
> - **B** — that is the lower boundary
> - **D** — that is the WIDTH of the interval

1.2 The MODAL interval is the one with ... (1)

A) the widest class **B) the highest frequency ✓**

C) the largest midpoint **D)** the middle position

> **Answer: B** — The mode is the most common value, so the modal interval is the fullest one.
>
> - **A** — class widths are usually equal, and width is not frequency
> - **C** — the largest midpoint means the highest values, not the most learners
> - **D** — that describes the interval containing the MEDIAN

1.3 Marks are grouped as $0–20$ $(f = 4)$, $20–40$ $(f = 9)$, $40–60$ $(f = 12)$ and $60–80$ $(f = 5)$. The MODAL interval is ... (1)

A) $60–80$ **B)** $20–40$

C) $0–20$ **D) $40–60$ ✓**

> **Answer: D** — A frequency of 12 is the highest of the four.
> - **A** — 5 is nearly the lowest frequency
> - **B** — 9 is the second-highest frequency, not the highest
> - **C** — 4 is the lowest frequency of all

1.4 For that grouped data, the total number of learners is ... (1)

A) 12 **B)** 26

C) 30 ✓ **D)** 20

> **Answer: C** — Add the four frequencies: $4 + 9 + 12 + 5 = 30$.
> - **A** — used the largest frequency instead of the total
> - **B** — left out one of the four frequencies
> - **D** — added the class widths instead of the frequencies

1.5 For that grouped data, the interval containing the MEDIAN is ... (1)

A) $0–20$ **B)** $60–80$

C) $20–40$ **D) $40–60$ ✓**

> **Answer: D** — The 15th and 16th values are the middle ones, and the running totals 4, 13 and 25 put both inside $40 – 60$.
> - **A** — the running total has only reached 4 by the end of $0–20$
> - **B** — the running total has already passed 16 before $60–80$ begins
> - **C** — the running total has only reached 13 by the end of $20 – 40$

1.6 For that grouped data, the ESTIMATED MEAN is ... (1)

A) 42 ✓ **B)** 40

C) $1\ 260$ **D)** 50

> **Answer: A** — $\Sigma fx = 40 + 270 + 600 + 350 = 1\ 260$, and $1\ 260 \div 30 = 42$.
> - **B** — averaged the four midpoints without weighting them by frequency
> - **C** — that is Σfx, before dividing by the 30 learners
> - **D** — took the midpoint of the modal interval instead of calculating the mean

1.7 The mean of grouped data is only an ESTIMATE because ... (1)

A) the intervals are unequal **B) the original values inside each interval are unknown** ✓

C) a mean is always approximate **D)** the frequencies are guesses

> **Answer: B** — Every value in an interval is treated as though it sat exactly on the midpoint.
> - **A** — these intervals are all equal in width, and it would still be an estimate
> - **C** — the mean of raw, ungrouped data is exact
> - **D** — the frequencies are counted exactly

1.8 The 25th percentile of a data set is the same as ... (1)

A) the median **B)** the mode

C) Q_3 **D)** **Q_1 ✓**

> **Answer: D** — The quartiles are the 25th, 50th and 75th percentiles.
>
> - **A** — the median is the 50th percentile
> - **B** — the mode is not a percentile at all
> - **C** — Q_3 is the 75th percentile

1.9 The SEMI-interquartile range is ... (1)

A) $\frac{Q_3 - Q_1}{2}$ ✓ **B)** $Q_3 \times Q_1$

C) $Q_3 - Q_1$ **D)** $\frac{Q_3 + Q_1}{2}$

> **Answer: A** — 'Semi' means half, so it is half of the interquartile range.
>
> - **B** — quartiles are never multiplied together
> - **C** — that is the interquartile range itself, before halving
> - **D** — that averages the two quartiles instead of halving their difference

1.10 If $Q_1 = 18$ and $Q_3 = 34$, the semi-interquartile range is ... (1)

A) 16 **B)** **8 ✓**

C) 26 **D)** 52

> **Answer: B** — $(34 - 18) \div 2 = 16 \div 2 = 8$.
>
> - **A** — that is the full interquartile range, before halving
> - **C** — averaged the two quartiles instead of halving their difference
> - **D** — added the two quartiles

ANSWER GRID — MARKING GUIDELINE

1.1	A	B	**C**	D
1.2	A	**B**	C	D
1.3	A	B	C	**D**
1.4	A	B	**C**	D
1.5	A	B	C	**D**
1.6	**A**	B	C	D
1.7	A	**B**	C	D
1.8	A	B	C	**D**
1.9	**A**	B	C	D
1.10	A	**B**	C	D

QUESTION 2 [16 MARKS]

The table below shows the times, in minutes, taken by 50 learners to complete a task.

Time (min)	$0 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 30$	$30 \leq t < 40$	$40 \leq t < 50$
Frequency	4	8	15	14	9

2.1 Complete a cumulative frequency table for the data. **(3)**

> 4; 12; 27; 41; 50 *(3)*

2.2 Calculate the estimated mean, using the midpoint of each interval. **(4)**

> $\Sigma fx = 20 + 120 + 375 + 490 + 405 = 1\,410$ *(3)*
>
> Estimated mean $= \frac{1\,410}{50} =$ **28,2 minutes** *(1)*

2.3 Determine the position of the UPPER QUARTILE, and the interval in which it lies. **(3)**

> Position $= \frac{75}{100} \times 50 = 37{,}5$ *(2)*
>
> The cumulative frequency is 27 at the end of the third interval and 41 at the end of the fourth, so Q_3 lies in $\mathbf{30 \leq t < 40}$ *(1)*

2.4 If $Q_1 = 21$ and $Q_3 = 37$, calculate the interquartile range and the semi-interquartile range, and state what the semi-interquartile range measures. **(3)**

> IQR $= 37 - 21 =$ **16** *(1)*
>
> Semi-IQR $= \frac{16}{2} =$ **8** *(1)*
>
> It measures the spread of the middle 50% of the data about the median *(1)*

2.5 Explain ONE advantage and ONE disadvantage of grouping data in this way. **(3)**

> Advantage: a large data set becomes easy to read and to summarise *(1)*
> Disadvantage: the individual values are lost *(1)*, so every calculated statistic is only an estimate *(1)*

TOTAL: 26 marks

This question paper consists of 2 questions.

© Danemathics — free to print and share for classroom use.