MARKING GUIDELINE

MARKS	DURATION	QUESTIONS
51	1 hour 20 minutes	4

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

Date: .....

MARK ____ / 51

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 The STANDARD DEVIATION measures ... (1)

A) the most common value
B) the middle value of the data
C) how spread out the data is about the mean ✓
D) the gap between the largest and smallest values

> **Answer: C** — It is the typical distance of a data value from the mean.
> - **A** — that is the MODE
> - **B** — that is the MEDIAN
> - **D** — that is the RANGE

1.2 If every value in a data set is increased by 5, the standard deviation ... (1)

A) stays the same ✓
B) is multiplied by 5
C) increases by 25
D) increases by 5

> **Answer: A** — Shifting every value moves the mean by the same amount, so the SPREAD does not change.
> - **B** — MULTIPLYING every value would change the spread; adding does not
> - **C** — 25 would belong to the variance, and only if the spread had changed
> - **D** — the MEAN increases by 5, but the spread does not

1.3 An OGIVE is a graph of ... (1)

A) frequency against the lower class boundary

B) cumulative frequency against frequency

C) cumulative frequency against the upper class boundary ✓

D) frequency against the class midpoint

> **Answer: C** — Each point sits at the upper boundary, because everything up to there has been counted.
>
> - **A** — the running total is complete only at the UPPER boundary
> - **B** — an ogive is plotted against the data values, not against frequency
> - **D** — that describes a FREQUENCY POLYGON

1.4 On an ogive of 80 values, the MEDIAN is read off at a cumulative frequency of ... (1)

A) 80

B) 40 ✓

C) 60

D) 20

> **Answer: B** — The median is the middle value, at half of 80.
>
> - **A** — that is the TOTAL number of values
> - **C** — that is where Q_3 is read off
> - **D** — that is where Q_1 is read off

1.5 A value is usually called an OUTLIER when it lies more than ... beyond the nearest quartile. (1)

A) half the range

B) $1 \times$ IQR

C) 2 standard deviations

D) $1{,}5 \times$ IQR ✓

> **Answer: D** — The usual convention marks a value as an outlier beyond $1{,}5$ interquartile ranges from Q_1 or Q_3.
>
> - **A** — half the range is not a standard outlier test
> - **B** — the standard multiplier is $1{,}5$, not 1
> - **C** — that is a different rule, and it is measured from the MEAN, not from a quartile

1.6 For 2 ; 4 ; 4 ; 4 ; 5 ; 5 ; 7 ; 9, the MEAN is ... (1)

A) 5 ✓

B) 40

C) $4{,}5$

D) 4

> **Answer: A** — The eight values total 40, and $40 \div 8 = 5$.
>
> - **B** — that is the TOTAL
> - **C** — that is the MEDIAN
> - **D** — that is the MODE

1.7 For that same data set, the STANDARD DEVIATION is ... (1)

A) $2{,}14$

B) 32

C) 4

D) 2 ✓

> **Answer: D** — The squared deviations total 32, and $\sqrt{32 \div 8} = \sqrt{4} = 2$.
>
> - **A** — divided by 7 instead of by 8
> - **B** — that is the total of the squared deviations
> - **C** — that is the VARIANCE, before the square root

1.8 On a box-and-whisker plot, a LONG right whisker suggests the data is ... (1)

A) symmetrical
B) skewed to the right ✓
C) bimodal
D) skewed to the left

> **Answer: B** — A long tail on the higher side is exactly what right, or positive, skew means.
> - **A** — symmetrical data has whiskers of similar length
> - **C** — a box plot shows nothing about the number of modes
> - **D** — a long LEFT whisker would show left skew

1.9 Data lies within ONE standard deviation of the mean when it lies between ... (1)

A) the mean and the mean plus σ
B) 0 and σ
C) the mean minus σ and the mean plus σ ✓
D) Q_1 and Q_3

> **Answer: C** — The interval reaches the same distance on BOTH sides of the mean.
> - **A** — the interval reaches below the mean as well as above it
> - **B** — the interval is measured from the MEAN, not from zero
> - **D** — that is the interquartile range, a different measure of spread

1.10 Two classes have the SAME mean, but class A has the larger standard deviation. This means ... (1)

A) class A had more learners
B) class A's marks are more spread out ✓
C) class A did better on average
D) class A necessarily had the highest single mark

> **Answer: B** — With equal means, the standard deviation is precisely what separates the two sets.
> - **A** — the standard deviation says nothing about how many learners there were
> - **C** — the two means are equal, so neither class did better on average
> - **D** — a wider spread usually means a higher maximum, but it is not guaranteed

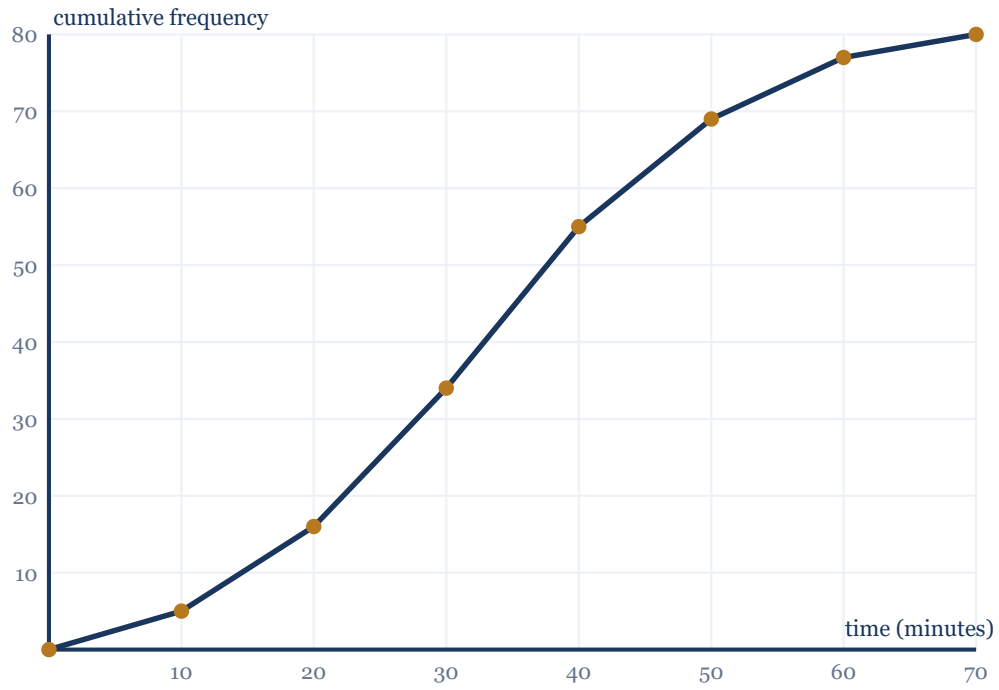
ANSWER GRID — MARKING GUIDELINE

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 [14 MARKS]

Eighty Grade 11 learners recorded the time, in minutes, that they spent on Mathematics homework on one evening. The results are summarised in the table below, and the ogive (cumulative frequency curve) for the same data is drawn beneath it.

Time (minutes)	$0 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 30$	$30 \leq t < 40$	$40 \leq t < 50$	$50 \leq t < 60$	$60 \leq t < 70$
Number of learners	5	11	18	21	14	8	3
Cumulative frequency	5	…	…	…	…	…	…



2.1 Complete the cumulative frequency row of the table. **(3)**

5 ; 16 ; 34 ; 55 ; 69 ; 77 ; 80 *(3)*
Each entry is the running total, so the last entry must be 80 — the size of the sample.

2.2 Use the ogive to estimate the median time spent on homework. Show, on the graph, how you obtained your answer. **(3)**

Read across from a cumulative frequency of $\frac{80}{2} = 40$ *(1)*
then down to the horizontal axis *(1)*
median ≈ **33 minutes** (accept 31 – 34) *(1)*

2.3 Use the ogive to estimate the interquartile range of the data. **(4)**

Q_1 is read at a cumulative frequency of 20: $Q_1 \approx 22{,}2$ min *(1)*
Q_3 is read at a cumulative frequency of 60: $Q_3 \approx 43{,}6$ min *(1)*
IQR $= Q_3 - Q_1 = 43{,}6 - 22{,}2$ *(1)*
≈ 21,4 minutes *(1)*
Readings off the graph: accept Q_1 from 21 to 23 and Q_3 from 43 to 45, with the IQR worked out from the learner's OWN two readings.

2.4 Estimate the number of learners who spent MORE than 40 minutes on homework, and hence write down the probability that a learner chosen at random from this group spent more than 40 minutes. **(4)**

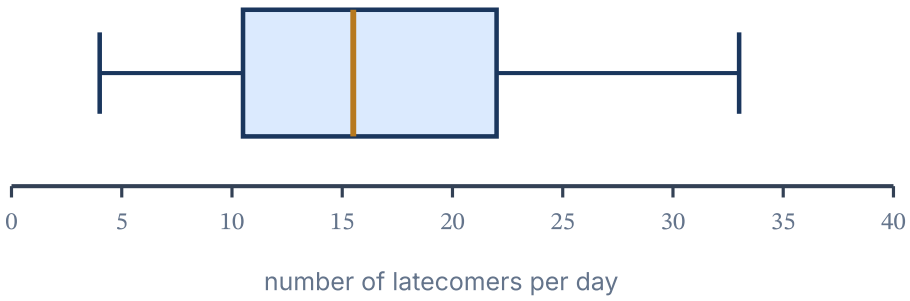
> At $t = 40$ the cumulative frequency is 55 *(1)*
> $80 - 55 = \mathbf{25}$ **learners** *(1)*
> $P = \frac{25}{80}$ *(1)*
> $= \mathbf{0,31}$ *(1)*

QUESTION 3 [17 MARKS]

The number of learners arriving late at a school was recorded on each of twelve days. The data, already arranged from smallest to largest, is:

4 ; 7 ; 9 ; 12 ; 12 ; 15 ; 16 ; 18 ; 21 ; 23 ; 28 ; 33

The box-and-whisker diagram below was drawn from this data.



3.1 Calculate the mean and the standard deviation of the data, correct to TWO decimal places. **(3)**

> mean $= 198 \div 12 = \mathbf{16,50}$ *(1)*
> $\sigma = \mathbf{8,24}$ *(2)*

3.2 Write down the five-number summary of the data and hence calculate the interquartile range. **(4)**

> Minimum $= 4$; $Q_1 = 10,5$; median $= 15,5$; $Q_3 = 22$; maximum $= 33$ *(3)*
> IQR $= 22 - 10,5 = \mathbf{11,5}$ *(1)*

3.3 Describe the skewness of this data, and justify your answer by referring BOTH to the box-and-whisker diagram and to the mean and the median. **(3)**

> The data is skewed to the RIGHT (positively skewed) *(1)*
> On the diagram the right whisker and the upper half of the box are the longer ones *(1)*
> and the mean (16,50) is greater than the median (15,5) *(1)*

3.4 A day is called a PROBLEM DAY if the number of latecomers is more than one standard deviation above the mean. Determine how many of the twelve days were problem days. **(3)**

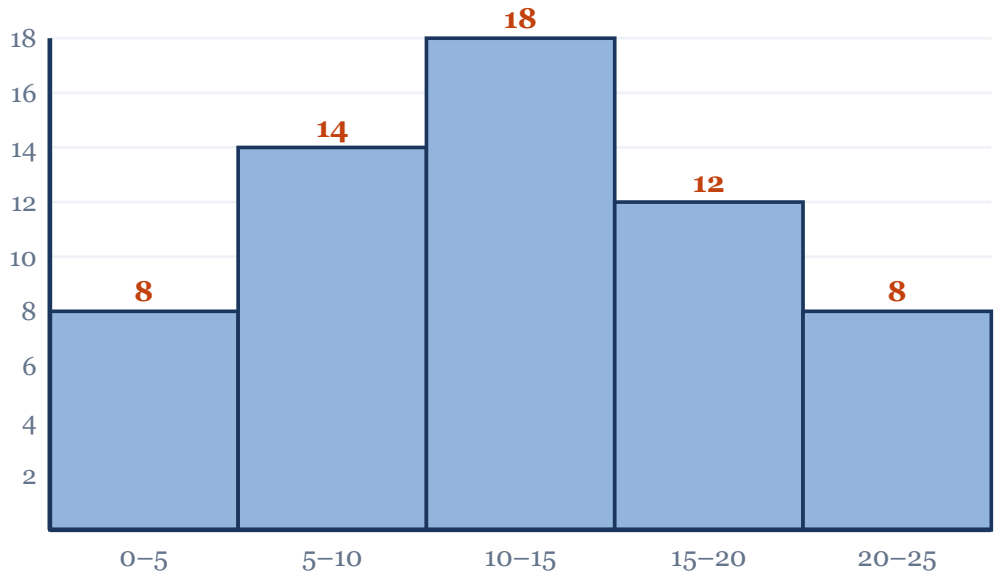
> mean $+ \sigma = 16,50 + 8,24$ *(1)*
> $= 24,74$ *(1)*
> Only 28 and 33 exceed 24,74, so there were **2 problem days** *(1)*

3.5 Use the $1{,}5 \times$ IQR rule to determine whether the largest value is an outlier. **(4)**

> Upper boundary $= Q_3 + 1{,}5 \times \text{IQR}$ *(1)*
> $= 22 + 1{,}5 \times 11{,}5$ *(1)*
> $= 39{,}25$ *(1)*
> $33 < 39{,}25$, so **33 is NOT an outlier** *(1)*

QUESTION 4 [10 MARKS]

The histogram below shows the mass, in kilograms, of 60 parcels handled at a depot on one morning.



mass (kg) — number of parcels on each bar

4.1 Use the midpoints of the intervals to estimate the mean mass of a parcel, correct to TWO decimal places. **(4)**

> Midpoints: 2,5 ; 7,5 ; 12,5 ; 17,5 ; 22,5 *(1)*
> $\Sigma(f \times x) = 8(2{,}5) + 14(7{,}5) + 18(12{,}5) + 12(17{,}5) + 8(22{,}5)$ *(1)*
> $= 740$ *(1)*
> estimated mean $= 740 \div 60 =$ **12,33 kg** *(1)*

4.2 Write down the modal class, and explain why the answer to the previous part can only ever be an ESTIMATE. **(3)**

> Modal class: $\mathbf{10 \leq m < 15}$ (the tallest bar, 18 parcels) *(1)*
> The individual masses are not known — only the interval each parcel falls into *(1)*
> so every parcel has to be treated as if it sat exactly at the midpoint of its interval *(1)*

4.3 Determine, with a reason, the interval in which the median mass lies. **(3)**

> $n = 60$, so the median is the average of the 30th and 31st values *(1)*
> Cumulative frequencies: 8 ; 22 ; 40 ; 52 ; 60 — the 30th and 31st values are both in the third interval *(1)*
> $\therefore$ the median lies in $\mathbf{10 \leq m < 15}$ *(1)*

TOTAL: 51 marks

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