Statistics: Five-Number Summary & Box Plots (Grade 10)

MARKING GUIDELINE

MARKS	DURATION	QUESTIONS
35	55 minutes	3

Name: _____ Class: _____

Date: _____

MARK ____ / 35

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 MEAN of 4 ; 7 ; 9 ; 11 ; 14 is ... (1)

A) 9 ✓ **B)** 10

C) 11,25 **D)** 45

Answer: A — $4 + 7 + 9 + 11 + 14 = 45$, and $45 \div 5 = 9$.

- **B** — that is the RANGE, $14 - 4$
- **C** — divided the total by 4 instead of by 5
- **D** — that is the TOTAL, not the mean

1.2 The MEDIAN of 3 ; 8 ; 5 ; 12 ; 7 is ... (1)

A) 5 **B)** 7,5

C) 12 **D)** 7 ✓

Answer: D — Order the data first: 3 ; 5 ; 7 ; 8 ; 12 — the middle value is 7.

- **A** — took the middle of the UNORDERED list
- **B** — averaged two middle values, but an odd-sized set has only ONE
- **C** — took the largest value instead of the middle one

1.3 The MODE of 2 ; 3 ; 3 ; 5 ; 7 ; 7 ; 7 ; 9 is ... (1)

A) 3 **B)** 5

C) 7 ✓ **D)** 5,375

Answer: C — 7 appears three times, more often than any other value.

- **A** — 3 appears only twice
- **B** — 5 sits near the middle — that is the median's job, not the mode's
- **D** — that is the MEAN, not the mode

1.4 For $2\ ;\ 5\ ;\ 6\ ;\ 9\ ;\ 11\ ;\ 14\ ;\ 18\ ;\ 20$, the lower quartile Q_1 is ... (1)

A) 5 **B) 5,5 ✓**

C) 6 **D)** 10

> **Answer: B** — The lower half is $2\ ;\ 5\ ;\ 6\ ;\ 9$, and its middle is $(5 + 6) \div 2 = 5{,}5$.
>
> - **A** — took only the lower of those two values
> - **C** — took only the upper of the two middle values of the lower half
> - **D** — that is the MEDIAN of the whole set

1.5 For that same data set, the INTERQUARTILE RANGE is ... (1)

A) 18 **B)** 5,5

C) 16 **D) 10,5 ✓**

> **Answer: D** — $\text{IQR} = Q_3 - Q_1 = 16 - 5{,}5 = 10{,}5$.
>
> - **A** — that is the RANGE, $20 - 2$
> - **B** — that is Q_1 on its own
> - **C** — that is Q_3 on its own

1.6 A five-number summary consists of ... (1)

A) mean, median, mode, range, IQR **B) minimum, Q_1, median, Q_3, maximum ✓**

C) the five largest values in the set **D)** minimum, mean, median, mode, maximum

> **Answer: B** — Those five numbers are exactly what a box-and-whisker plot is drawn from.
>
> - **A** — those are five statistics, but not the five-number summary
> - **C** — a summary describes the whole set, not only its largest values
> - **D** — the mean and the mode are not part of it — the QUARTILES are

1.7 On a box-and-whisker plot, the length of the BOX shows ... (1)

A) the interquartile range ✓ **B)** the range

C) the median **D)** the mean

> **Answer: A** — The box runs from Q_1 to Q_3, so its length is $Q_3 - Q_1$.
>
> - **B** — the range is the whole plot, whisker tip to whisker tip
> - **C** — the median is the LINE inside the box, not the box's length
> - **D** — the mean does not appear on a box plot at all

1.8 For a data set with $Q_1 = 14$, $Q_2 = 21$ and $Q_3 = 32$, the interquartile range is ... (1)

A) 7 **B)** 11

C) 18 ✓ **D)** 21

> **Answer: C** — $\text{IQR} = Q_3 - Q_1 = 32 - 14 = 18$.
>
> - **A** — that is $Q_2 - Q_1$, only half of the box
> - **B** — that is $Q_3 - Q_2$, the upper half of the box
> - **D** — that is the median, not a measure of spread

1.9 Salaries are R8 000 ; R9 000 ; R9 500 ; R10 000 ; R120 000. The better measure of the centre is ... (1)

A) the range, because it shows the spread
B) the median, because R120 000 is an outlier ✓
C) the mean, because it uses every value
D) the mode, because it is the most common value

> **Answer: B** — The mean is R31 300, which is higher than four of the five salaries.
> - **A** — the range measures SPREAD, not the centre
> - **C** — using every value is exactly what lets one outlier drag the mean up
> - **D** — no salary repeats, so there is no mode

1.10 Adding one more value, equal to the mean, leaves the mean ... (1)

A) impossible to determine
B) larger
C) unchanged ✓
D) smaller

> **Answer: C** — The total and the number of values grow in exactly matching proportion.
> - **A** — it can be determined exactly — the mean stays the same
> - **B** — the new value is not above the mean, so it cannot pull the mean up
> - **D** — the new value is not below the mean, so it cannot pull the mean down

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 [13 MARKS]

Thirteen people were asked how many kilometres they drive in a day. The results are given below.

5 10 10 12 15 20 22 35 35 40 60 90 105

2.1 Calculate the mean of the data, correct to TWO decimal places. (3)

> Total = 459 *(1)*
> $\frac{459}{13}$ *(1)* = **35,31 km** *(1)*

2.2 Write down the five-number summary for the data. (4)

> Minimum 5 and maximum 105 *(1)*
> $Q_1 = 11$ *(1)*
> median = 22 *(1)*
> $Q_3 = 50$ *(1)*

2.3 Calculate the interquartile range, and explain what it measures that the RANGE does not. **(3)**

> IQR $= 50 - 11$ *(1)*
> $= \mathbf{39}$ *(1)*
> The IQR measures the spread of the MIDDLE half of the data, so unlike the range it is not stretched by the largest value, 105 *(1)*

2.4 The mean is much larger than the median. State which of the two better describes a typical value here, and give TWO reasons. **(3)**

> The **median** *(1)*
> The few large values (90 and 105) pull the mean up *(1)*
> The median depends only on the middle position, so extreme values do not move it *(1)*

QUESTION 3 [12 MARKS]

The Mathematics marks of 50 Grade 10 learners were recorded.

Marks	$0 \leq x < 10$	$10 \leq x < 20$	$20 \leq x < 30$	$30 \leq x < 40$	$40 \leq x < 50$
Frequency	4	8	15	14	9

3.1 Write down the midpoint of each interval, and explain why the midpoints have to be used instead of the actual marks. **(3)**

> 5; 15; 25; 35; 45 *(2)*
> The individual marks are not known — the table records only which interval each learner falls into, so every learner is treated as if the mark sat at the middle of the interval *(1)*

3.2 Calculate the ESTIMATED mean of the data. **(4)**

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

3.3 Write down the modal interval, and determine the interval in which the median lies. **(3)**

> Modal interval: $\mathbf{20 \leq x < 30}$ *(1)*
> The median is between the 25th and 26th value; the cumulative frequency reaches 27 at the end of the third interval *(1)*
> Median interval: $\mathbf{20 \leq x < 30}$ *(1)*

3.4 Explain why the mean calculated above is only an ESTIMATE. **(2)**

> The individual marks are not known once the data is grouped *(1)*; every learner in an interval is assumed to have scored the midpoint *(1)*

TOTAL: 35 marks

This question paper consists of 3 questions.

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