

STATISTICS

Discrete data: information collected by counting

Continuous data: information collected by measurement, e.g.
lengths, heights, time, etc.

A. Ungrouped data (arrange in order of size.)

1. **Mode:** score that occurs most often (number with greatest frequency)

2. **Median:** middle number in a set of ranked data

3. **Mean (average):** $\bar{x} = \frac{\text{total value of numbers listed}}{\text{total frequency}}$ $\bar{x} = \frac{\sum x}{n}$

4. **Range:** largest value - smallest value

5. **Quartiles:** divide data spread into quarters

6. **Percentiles:** divide data spread into **one hundredths**

position of 30th percentile: e.g. $P_{30} = \frac{30}{100}(n+1)$

7. **Lower quartile Q_1 :** 25th percentile: position = $\frac{1}{4}(n+1)$

8. **Median Q_2 :** 50th percentile: position = $\frac{1}{2}(n+1)$

9. **Upper quartile Q_3 :** 75th percentile: position = $\frac{3}{4}(n+1)$

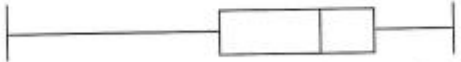
10. **Interquartile range IQR:** $Q_3 - Q_1$

11. **Semi-interquartile range:** $\frac{Q_3 - Q_1}{2}$

12. **Five number summary:** sample size (min.; Q_1 ; M; Q_3 ; max.)

13. **Box and whisker diagram:**

13.1 Symmetrical data min. value |  | max. value

13.2 Skewed data | 
skewed to the left (data is spread more to the left of the median)
[NB. Draw diagram to scale on a number line.]

14. **Modal class:** class with highest frequency

15. **Ogive:** cumulative frequency polygon

16. **Outliers:** data values less than $[Q_1 - 1,5 \times IQR]$
 data values greater than $[Q_3 + 1,5 \times IQR]$

B. Grouped data

17. Σ : sum of
 n : total number of scores
 x : each score in sample
 x_1 : midpoint of interval
 f : frequency of score

18. **Mean:** $\bar{x} = \frac{\sum f \cdot x_1}{n}$

19. **Variance:** $\frac{\sum (x_i - \bar{x})^2}{n}$

20. **Standard deviation:** $\sqrt{\text{variance}}$

Example 1

Data	f	cum. f	x_1	$f \cdot x_1$
$10 \leq x < 20$	2	2	15	30
$20 \leq x < 30$	5	7	25	125
$30 \leq x < 40$	18	25	35	630
$40 \leq x < 50$	20	45	45	900
$50 \leq x < 60$	15	60	55	825
				2510

Mean: $\bar{x} = \frac{\sum f \cdot x_1}{n} = \frac{2510}{60} = 41,83$

Example 2

Data(x)	$x - \bar{x}$	$(x - \bar{x})^2$
8	-3	9
10	-1	1
12	1	1
14	3	9
11	0	0
55		20

1. **mean:** $\bar{x} = \frac{55}{5} = 11$

2. **Variance:** $\frac{\sum (x - \bar{x})^2}{n} = \frac{20}{5} = 4$

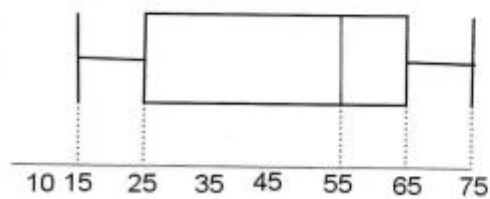
3. **Standard dev. (σ):** $\sqrt{4} = 2$

MAKE SURE THAT YOU CAN DETERMINE THE STANDARD DEVIATION SD (σ), BY USING YOUR CALCULATOR. A high standard deviation indicates that the data values are spread out and have a large range.

Worksheet 7 A

Given: 15; 24; 11; 22; 8; 78; 33;
31; 83; 15; 26; 22; 36;
21; 36; 22

1. Determine the
 - 1.1 mean;
 - 1.2 mode;
 - 1.3 median;
 - 1.4 lower quartile Q_1 ;
 - 1.5 upper quartile Q_3 ;
 - 1.6 range ;
 - 1.7 inter-quartile range (IQR);
 - 1.8 semi- IQR;
 - 1.9 outliers.
2. The coins in 10 money bags were counted. The result was:
20; $2x$; 15; 30; 32; 42;
 x ; 21; 22; 18
 - 2.1 The average number of coins per bag is 23. Calculate the value of x to the nearest integer.
 - 2.2 Use the information to draw a box and whisker diagram.
- 3.



- 3.1 Which percentage of the data lies between 15 and 55?
- 3.2 Calculate the difference between the median and the third quartile.

Worksheet 7 A

- 3.3 Determine the range of the data.
- 3.4 Comment on the skewedness of the data.
- 3.5 What percentage of the data is more than 65?
4. The average of 3 friends' mass is 95 kg. Frans's mass is 100 kg. Nico's mass is 87 kg. Determine Jaco's mass.
5. Given:
50 customers were asked to record the amount of money spent on the 8th of February 2012. The following table shows the purchases:

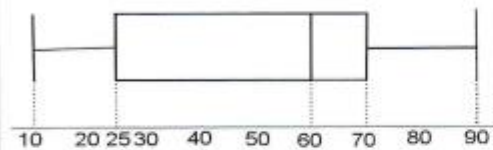
Groceries purchased (R)	Number of clients
$0 \leq x < 50$	1
$50 \leq x < 100$	5
$100 \leq x < 150$	6
$150 \leq x < 200$	18
$200 \leq x < 250$	14
$250 \leq x < 300$	6

- 5.1 Determine the
 - (a) modal class;
 - (b) median class;
 - (c) mean amount that was spent on groceries.
- 5.2 Draw a cumulative frequency graph of the data.
- 5.3 Use the graph to find the median, lower and upper quartile.

Worksheet 7 C

3. Two schools, A and B, compare their "AdMaths" students' marks in an examination. School A provides learners' information with a box and whisker diagram. School B provides only the marks. There are 12 students in each school who take "AdMaths".

School A:



School B:

30; 43; 84; 68; 45; 90; 86;
60; 74; 70; 78; 56

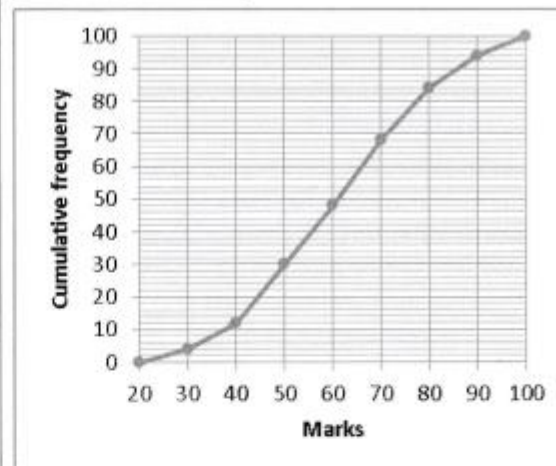
- 3.1 How would you compare the two schools' marks?
3.2 Show with reasons and the necessary calculations, which school has the best results.
3.3 Determine the interquartile range of school A.
3.4 Give the interval of points for the top 3 students in class A.

4.

The ogive represents the examination marks of 100 gr. 11 learners.

- 4.1 Use the ogive to complete the given frequency table.

Worksheet 7 C



Marks	Cum f	f
$20 \leq x < 30$		4
$30 \leq x < 40$	12	
$40 \leq x < 50$		
$50 \leq x < 60$		18
$60 \leq x < 70$	68	
$70 \leq x < 80$		
$80 \leq x < 90$	94	10
$90 \leq x < 100$	100	

- 4.2 Determine how many learners achieved 50% or more in the examination.
5. 2 hockey teams' goal keepers are compared with respect to goals prevented per game. Their results are as follows:

Sipho	6	8	7	3	8	5
Johan	2	12	2	10	3	7

- 5.1 Calculate the average number of goals prevented for every goal keeper.
5.2 Calculate the standard deviation of the data for each player.
5.3 Compare the 2 players using norms of central tendency and measures of dispersion.

Worksheet 7 C

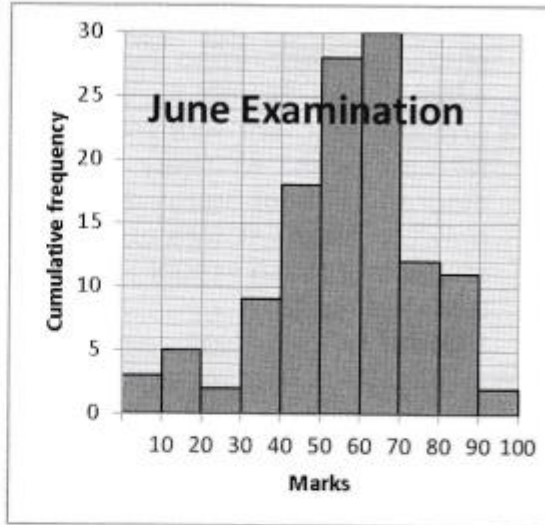
6. The municipality purchases a truck at the beginning of the year for R2 000 000. The truck depreciates at a rate of 12% per annum. The table gives the value of the truck at the end of each of the seven years of its serviceable life.

Year end	Truck value (R m)
1	1,760
2	1,549
3	1,363
4	1,199
5	1,055
6	0,929
7	0,817

- 6.1 Draw a scatter plot to represent the above data.
- 6.2 Describe the nature of the number pattern in the second column. Give a reason for your answer.
- 6.3 Draw a curve of best fit.
- 6.4 It is suspected that the curve of best fit has an equation of the form $W=ab^x$. Using the information you have obtained in this question, give the equation of the curve.
7. Examination marks obtained by 120 grade 11 learners in the June examinations are indicated in the following histogram:

Worksheet 7 C

7.



- 7.1 Complete a cumulative frequency table for the data above.

Marks	f	Cum. f
$0 \leq x < 10$	3	
$10 \leq x < 20$	5	
	2	
$30 \leq x < 40$		
$40 \leq x < 50$		
$60 \leq x < 70$		
$70 \leq x < 80$		
$90 \leq x < 100$		

- 7.2 Draw an ogive of the data.
- 7.3 Use the ogive and determine:
- 7.3.1 how many learners achieved more than 75%;
- 7.3.2 the median.