BELGIUM CAMPUS iTversity

It's the way we're wired

GRADE 11 MATHS

Matildah Chiruka

COURSE CONTENT

- Skewness of data
- Variance
- Curve fitting
- Histograms
- Ogives

SKEWNESS OF DATA

- Symmetrical is when then the median=mean or approximately equal
- Skewed to the right is when the longer tail is to the right median is less than mean
- Skewed to the left is when the longer tail is to the left median is greater than the mean
- To determine skewness of data different diagrams can be used e.g. normal curves, frequency polygons, histograms and box and whisker.

For each of the following data sets, compute the mean and all the quartiles. Round your answers to one decimal place.

a) $-3{,}4$; $-3{,}1$; $-6{,}1$; $-1{,}5$; $-7{,}8$; $-3{,}4$; $-2{,}7$; $-6{,}2$

b) -6 ; -99 ; 90 ; 81 ; 13 ; -85 ; -60 ; 65 ; -49

c) 7 ; 45 ; 11 ; 3 ; 9 ; 35 ; 31 ; 7 ; 16 ; 40 ; 12 ; 6

a) Mean:

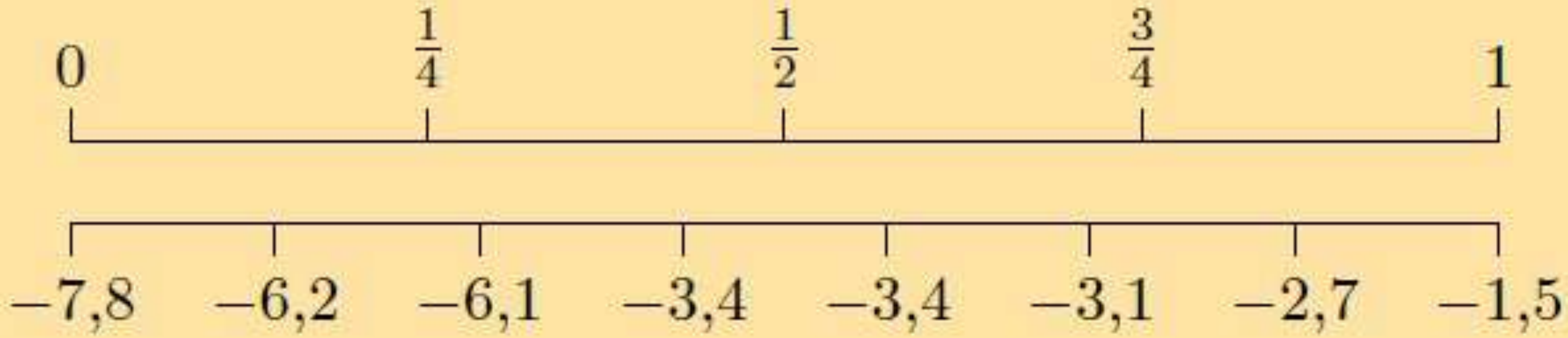
$$\overline{x} = \frac{(-3{,}4) + (-3{,}1) + (-6{,}1) + (-1{,}5) + (-7{,}8) + (-3{,}4) + (-2{,}7) + (-6{,}2)}{8}$$

$$\approx -4{,}3$$

To compute the quartiles, we order the data:

$-7{,}8$; $-6{,}2$; $-6{,}1$; $-3{,}4$; $-3{,}4$; $-3{,}1$; $-2{,}7$; $-1{,}5$

We use the diagram below to find at or between which values the quartiles lie.



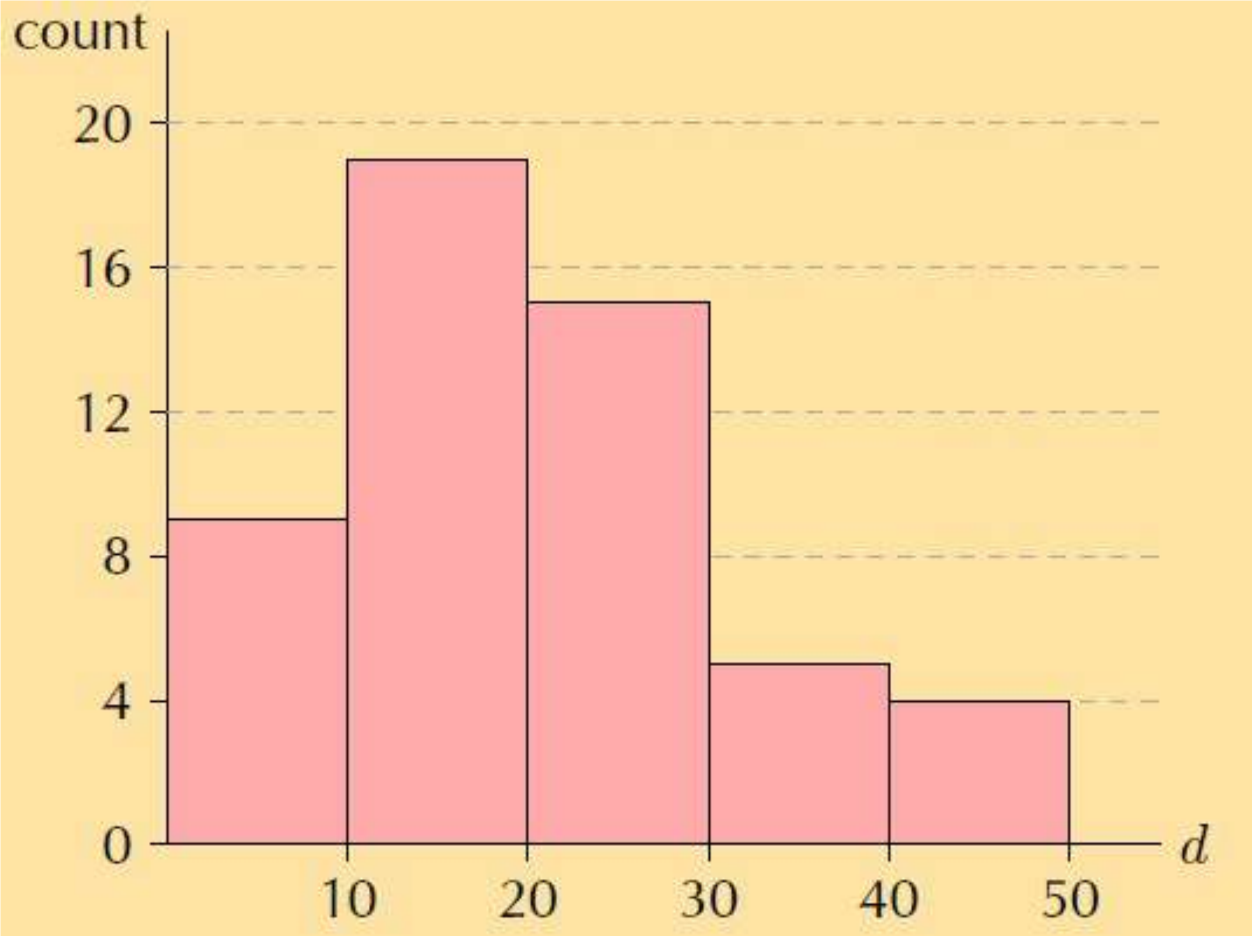
For the first quartile the position is between the second and third values. The second value is $-6{,}2$ and the third value is $-6{,}1$, which means that the first quartile is $\frac{-6{,}2-6{,}1}{2} = -6{,}15$.

For the median (second quartile) the position is halfway between the fourth and fifth values. Since both these values are $-3{,}4$, the median is $-3{,}4$.

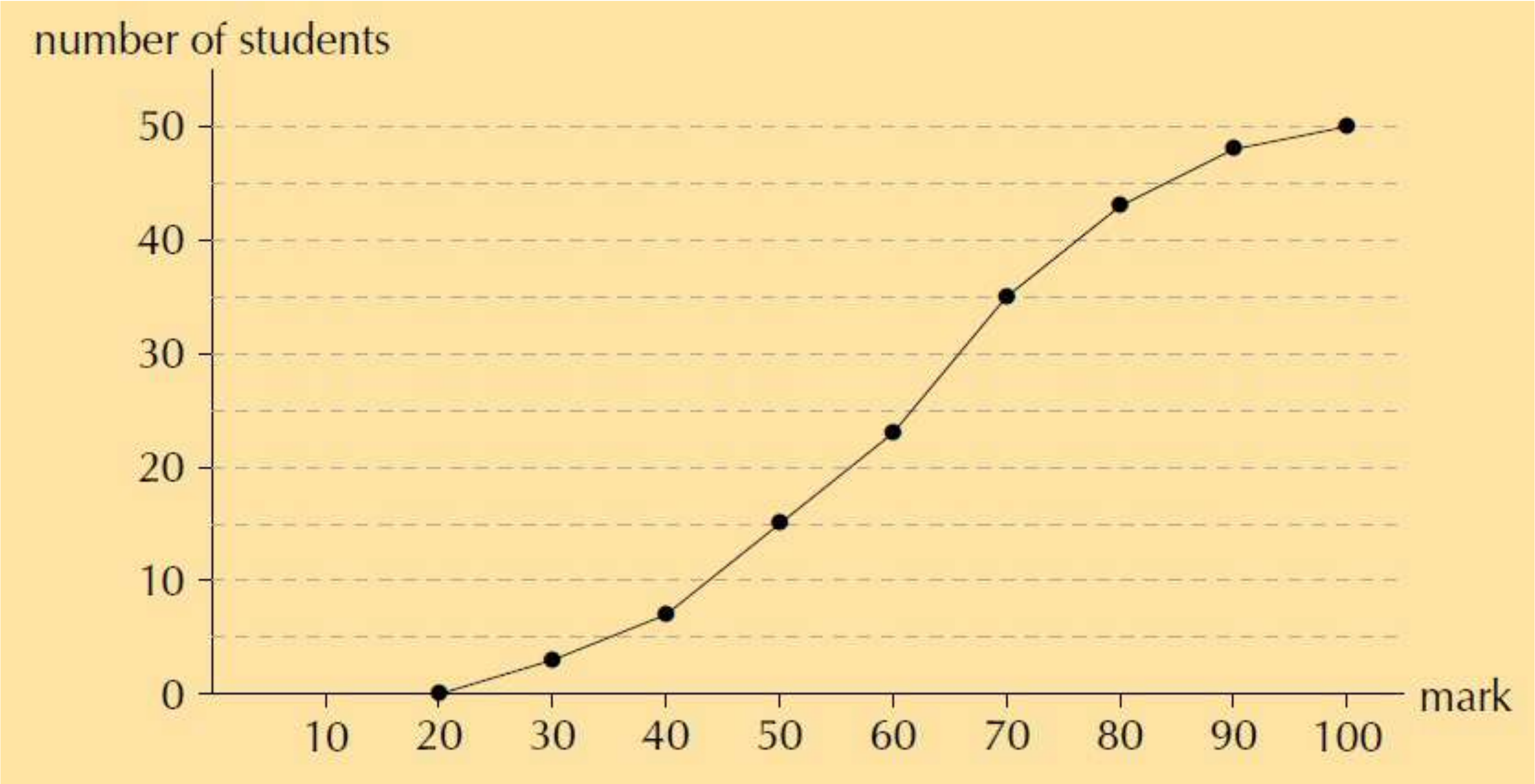
For the third quartile the position is between the sixth and seventh values. Therefore the third quartile is $-2{,}9$.

In a traffic survey, a random sample of 50 motorists were asked the distance they drove to work daily. The results of the survey are shown in the table below. Draw a histogram to represent the data.

d (km)	$0 < d \leq 10$	$10 < d \leq 20$	$20 < d \leq 30$	$30 < d \leq 40$	$40 < d \leq 50$
f	9	19	15	5	4


count
20
16
12
8
4
0
10
20
30
40
50
d

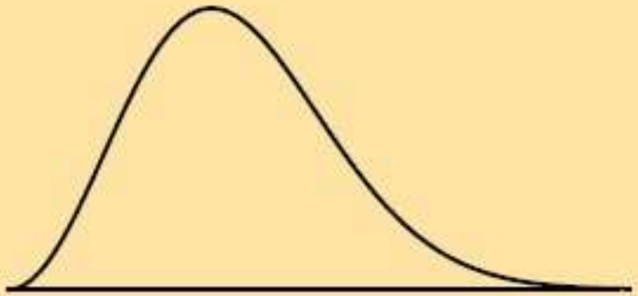
OGIVE



number of students
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
mark

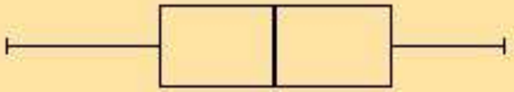
EXAMPLES

a) A data set with this distribution:



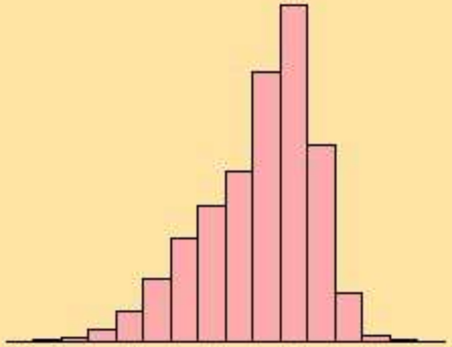
Solution: skewed right

b) A data set with this box and whisker plot:



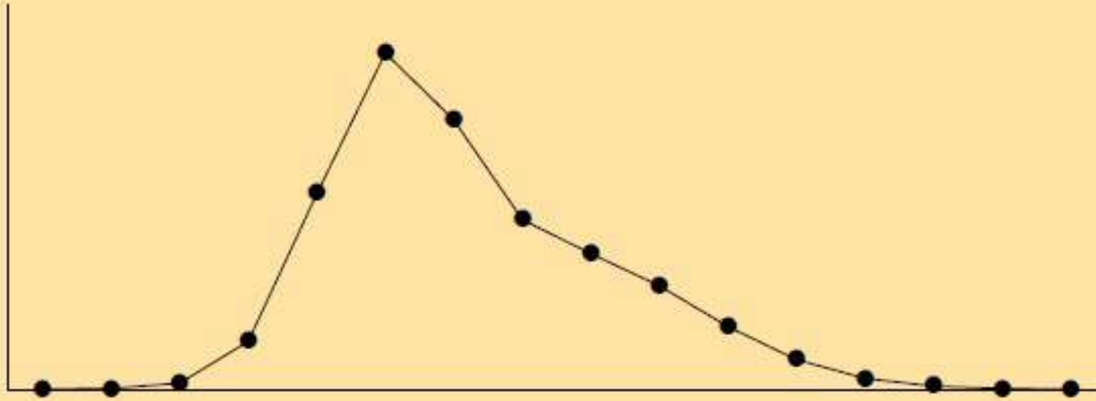
Solution: symmetric

c) A data set with this histogram:



Solution: skewed left

d) A data set with this frequency polygon:



VARIANCE EXAMPLES

{9; 5; 1; 3; 3; 5; 7; 4; 10; 8}

Solution:

The formula for the mean is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\therefore \bar{x} = \frac{55}{10}$$

$$= 5{,}5$$

The formula for the variance is

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}$$

We first subtract the mean from each data point and then square the result.

x_i	9	5	1	3	3	5	7	4	10	8
$x_i - \overline{x}$	3,5	−0,5	−4,5	−2,5	−2,5	−0,5	1,5	−1,5	4,5	2,5
$(x_i - \overline{x})^2$	12,25	0,25	20,25	6,25	6,25	0,25	2,25	2,25	20,25	6,25

EXERCISE

The variance is the sum of the last row in this table divided by 10, so $\sigma^2 = \frac{76,5}{10} = 7,65$. The standard deviation is the square root of the variance, therefore $\sigma = \sqrt{7,65} = \pm 2,77$.

The interval containing all values that are one standard deviation from the mean is $[5,5 - 2,77; 5,5 + 2,77] = [2,73; 8,27]$. We are asked how many values are **within** than one standard deviation from the mean, meaning **inside** the interval. There are 7 values from the data set within the interval, which is $\frac{7}{10} \times 100 = 70\%$ of the data points.

1. Draw a histogram, frequency polygon and ogive of the following data set. To count the data, use intervals with a width of 1, starting from 0.

   0,4 ; 3,1 ; 1,1 ; 2,8 ; 1,5 ; 1,3 ; 2,8 ; 3,1 ; 1,8 ; 1,3 ;

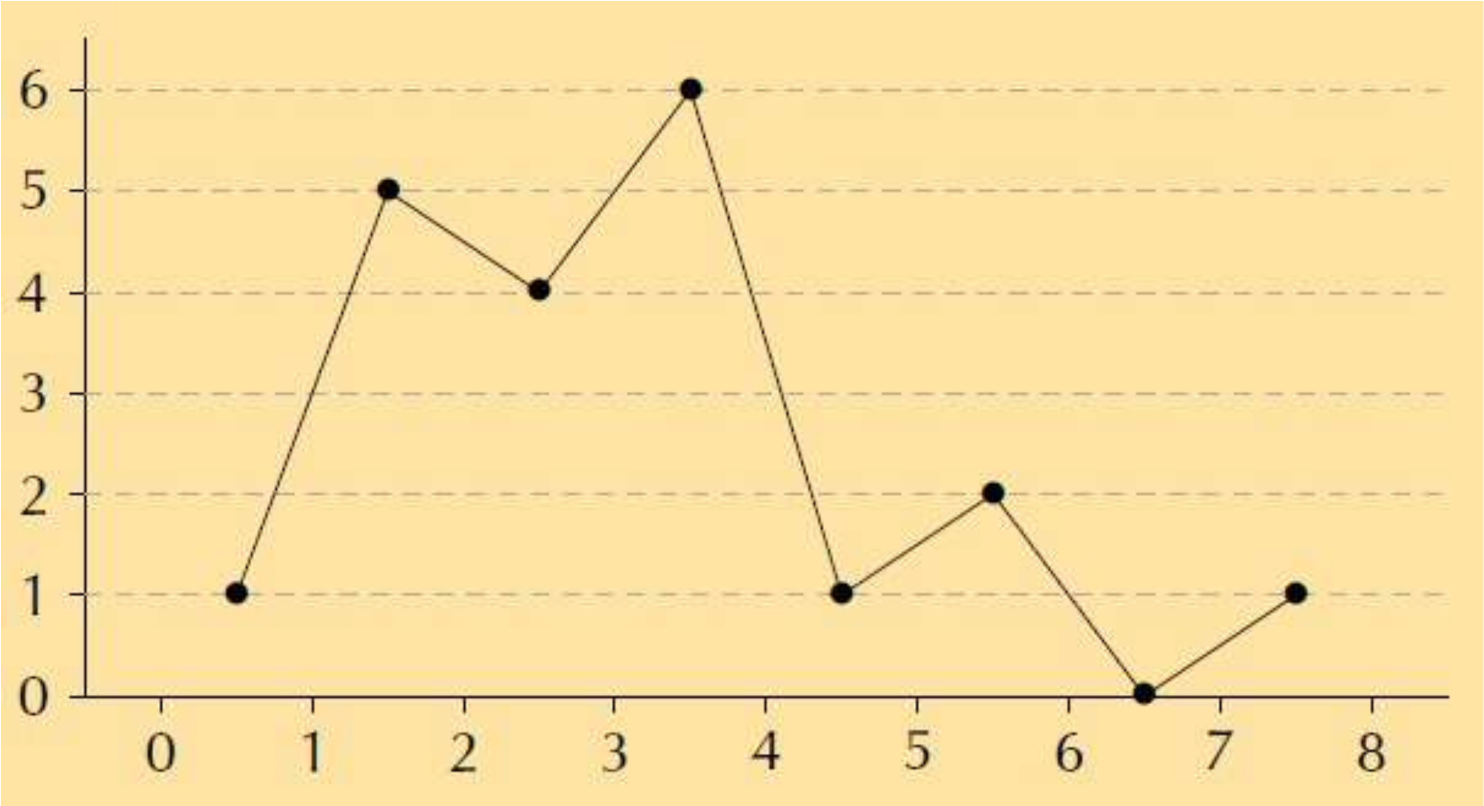
   2,6 ; 3,7 ; 3,3 ; 5,7 ; 3,7 ; 7,4 ; 4,6 ; 2,4 ; 3,5 ; 5,3

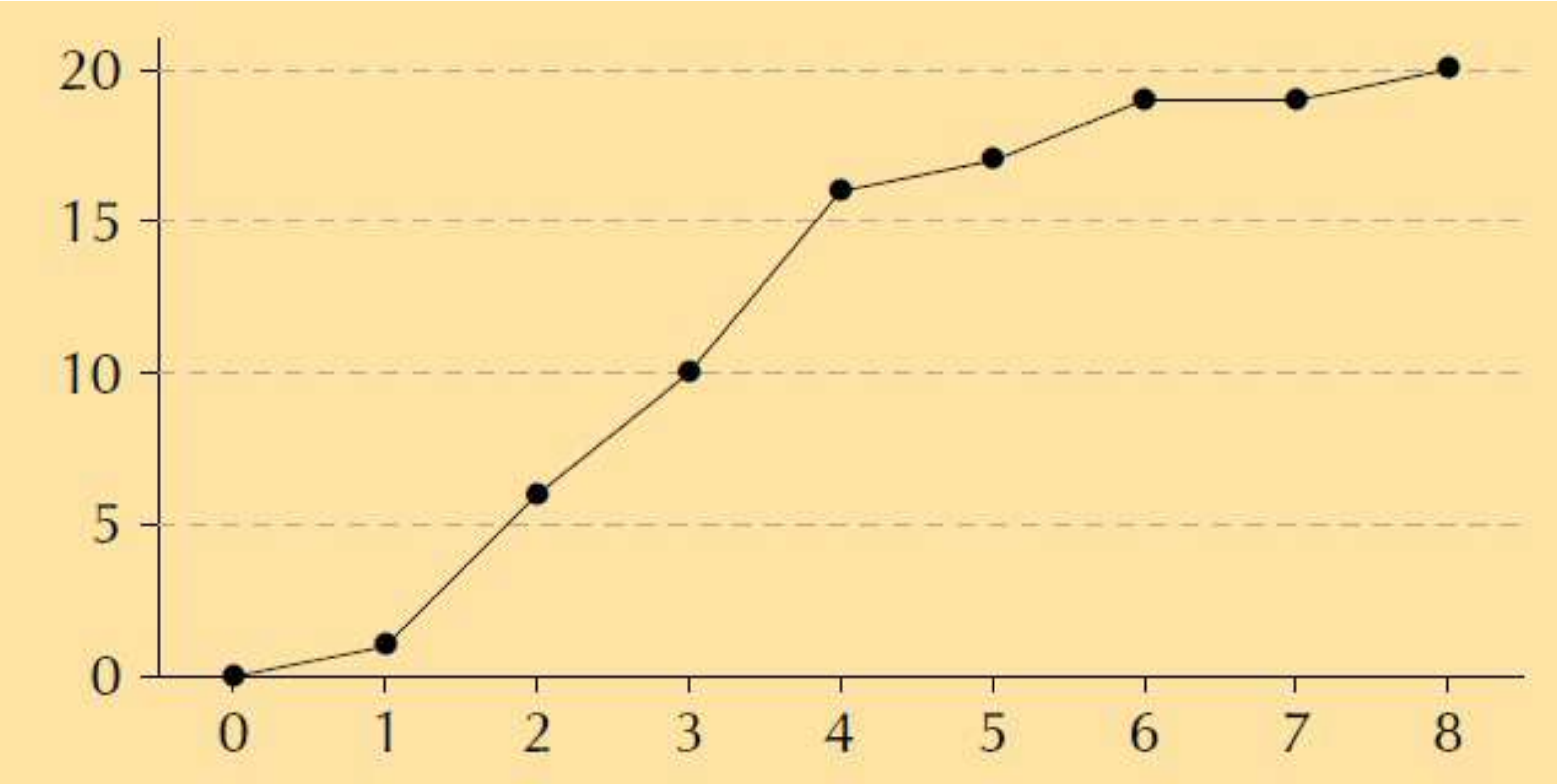
   Solution:

   We first organise the data into a table using an interval width of 1, showing the count in each interval as well as the cumulative count across intervals.

Interval	$[0;1)$	$[1;2)$	$[2;3)$	$[3;4)$	$[4;5)$	$[5;6)$	$[6;7)$	$[7;8)$
Count	1	5	4	6	1	2	0	1
Cumulative	1	6	10	16	17	19	19	20

From the table above we can draw the histogram, frequency polygon and ogive.


6
5
4
3
2
1
0
0 1 2 3 4 5 6 7 8



20
15
10
5
0
0
1
2
3
4
5
6
7
8

2. Draw a box and whisker diagram of the following data set and explain whether it is symmetric, skewed right or skewed left.

$-4{,}1$; $-1{,}1$; -1 ; $-1{,}2$; $-1{,}5$; $-3{,}2$; -4 ; $-1{,}9$; -4 ;

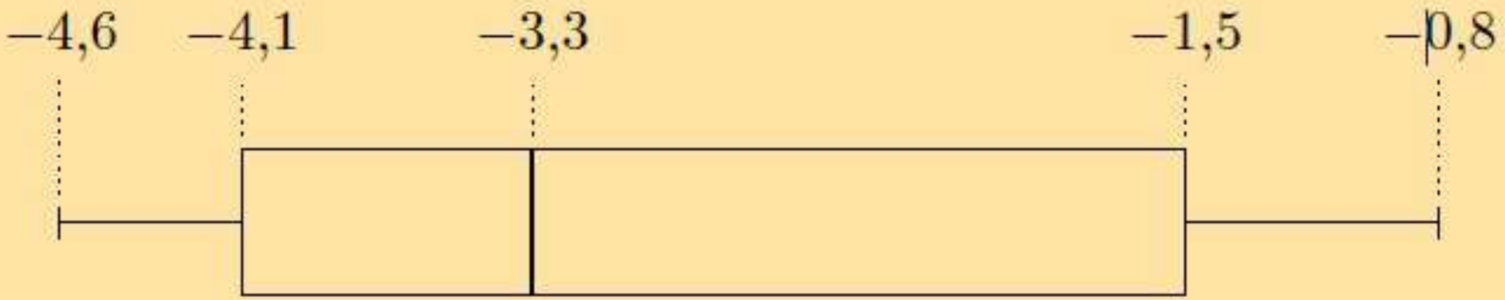
$-0{,}8$; $-3{,}3$; $-4{,}5$; $-2{,}5$; $-4{,}4$; $-4{,}6$; $-4{,}4$; $-3{,}3$

Solution:

The statistics of the data set are

- minimum: $-4{,}6$;
- first quartile: $-4{,}1$;
- median: $-3{,}3$;
- third quartile: $-1{,}5$;
- maximum: $-0{,}8$.

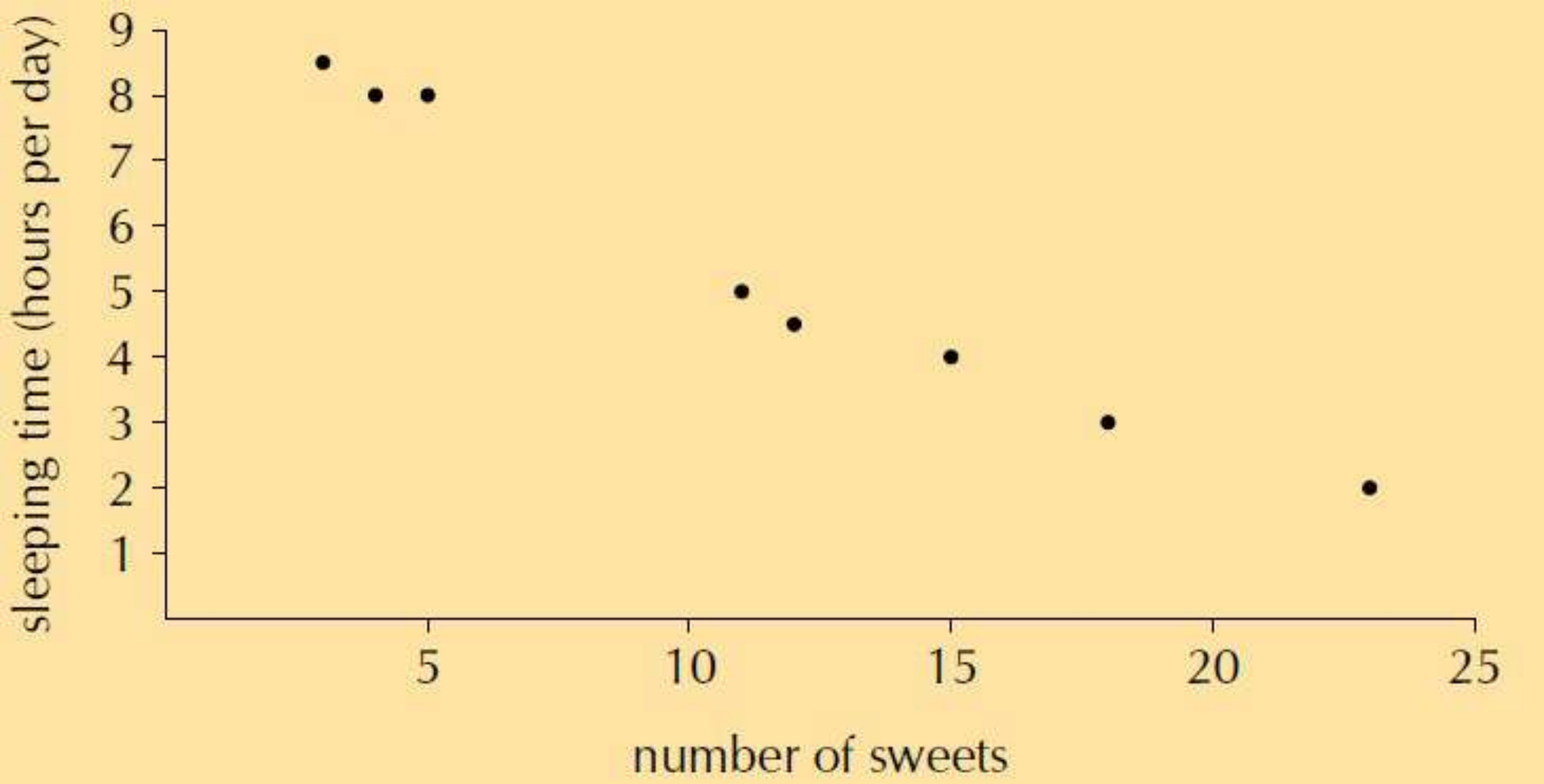
From this we can draw the box-and-whisker plot as follows.



Since the median is closer to the first quartile than the third quartile, the data set is skewed right.

3. Eight children's sweet consumption and sleeping habits were recorded. The data are given in the following table and scatter plot.

Number of sweets per week	15	12	5	3	18	23	11	4
Average sleeping time (hours per day)	4	4,5	8	8,5	3	2	5	8



sleeping time (hours per day)
9
8
7
6
5
4
3
2
1
5
10
15
20
25
number of sweets

a) What is the mean and standard deviation of the number of sweets eaten per day?

b) What is the mean and standard deviation of the number of hours slept per day?

c) Make a list of all the outliers in the data set.

Solution:

a) Mean $= 11\frac{3}{8}$. Standard deviation $= 6{,}69$.

b) Mean $= 5\frac{3}{8}$. Standard deviation $= 2{,}33$.

c) There are no outliers.

4. The monthly incomes of eight teachers are as follows:

R 10 050; R 14 300; R 9800; R 15 000; R 12 140; R 13 800; R 11 990; R 12 900.

a) What is the mean and standard deviation of their incomes?

b) How many of the salaries are less than one standard deviation away from the mean?

BELGIUM CAMPUS iTversity
It's the way we're wired

c) If each teacher gets a bonus of R 500 added to their pay what is the new mean and standard deviation?

d) If each teacher gets a bonus of 10% on their salary what is the new mean and standard deviation?

e) Determine for both of the above, how many salaries are less than one standard deviation away from the mean.

f) Using the above information work out which bonus is more beneficial financially for the teachers.

Solution:

a) Mean = R 12 497,50. Standard deviation = R 1768,55.

b) All salaries within the range (10 728,95 ; 14 266,05) are less than one standard deviation away from the mean. There are 4 salaries inside this range.

c) Since the increase in each salary is the same absolute amount, the mean simply increases by the bonus. The standard deviation does not change since every value is increased by exactly the same amount. Mean = R 12 997,50. Standard deviation = R 1768,55.

d) With a relative increase, the mean and standard deviation are both multiplied by the same factor. With an increase of 10% the factor is 1,1. Mean = R 13 747,25. Standard deviation = R 1945,41.

e) Adding a constant amount or multiplying by a constant factor (that is, applying a linear transformation) does not change the number of values that lie within one standard deviation from the mean. Therefore the answer is still 4.

f) Since the mean is greater in the second case it means that, on average, the teachers are getting better salaries when the increase is 10%.

QUESTIONS GRADE 11

BELGIUM CAMPUS iTversity
It's the way we're wired

The table below shows the number of cans of food collected by 9 classes during a charity drive.

5	8	15	20	25	27	31	36	75

1.1 Calculate the range of the data.

1.2 Calculate the standard deviation of the data.

1.3 Determine the median of the data.

1.4 Determine the interquartile range of the data.

1.5 Use the number line provided in the ANSWER BOOK to draw a box and whisker diagram for the data above.

1.6 Describe the skewness of the data.

1.7 Identify outliers, if any exist, for the above data.

The table below shows the time (in minutes) that 200 learners spent on their cellphones during a school day.

TIME SPENT (IN MINUTES)	FREQUENCY
$95 < x \leq 105$	15
$105 < x \leq 115$	27
$115 < x \leq 125$	43
$125 < x \leq 135$	52
$135 < x \leq 145$	28
$145 < x \leq 155$	21
$155 < x \leq 165$	10
$165 < x \leq 175$	4

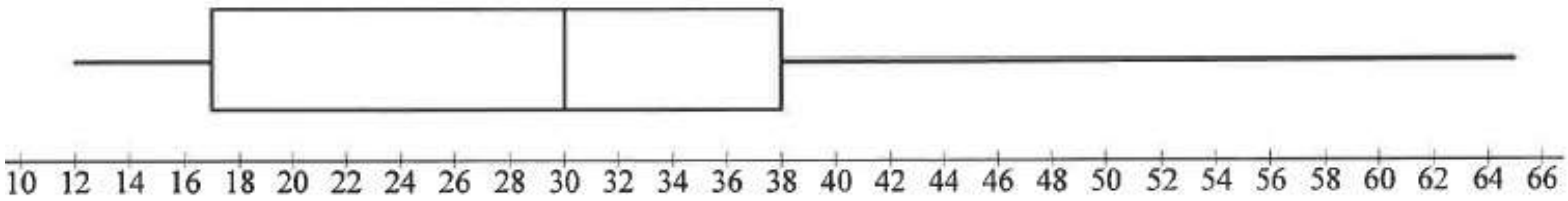
2.1 Complete the cumulative frequency column in the table provided in the ANSWER BOOK. (2)

2.2 Draw a cumulative frequency graph (ogive) of the data on the grid provided. (3)

2.3 Use the cumulative frequency graph to determine the value of the lower quartile. (2)

2.4 Determine, from the cumulative frequency graph, the number of learners who used their cellphones for more than 140 minutes. (2)

1.1 Mr Brown conducted a survey on the amount of airtime (in rands) EACH student had on his or her cellphone. He summarised the data in the box and whisker diagram below.



1.1.1 Write down the five-number summary of the data. (2)

1.1.2 Determine the interquartile range. (1)

1.1.3 Comment on the skewness of the data. (1)

1.2 A group of 13 students indicated how long it took (in hours) before their cellphone batteries required recharging. The information is given in the table below.

5	8	10	17	20	29	32	48	50	50	63	y	107

1.2.1 Calculate the value of y if the mean for this data set is 41. (2)

1.2.2 If $y = 94$, calculate the standard deviation of the data. (1)

1.2.3 The mean time before another group of 6 students needed to recharge the batteries of their cellphones was 18 hours. Combine these groups and calculate the overall mean time needed for these two groups to recharge the batteries of their cellphones. (3)

A student conducted a survey among his friends and relatives to determine the relationship between the age of a person and the number of marketing phone calls he or she received within one month. The information is given in the table below.

2.1 Complete the frequency and cumulative frequency columns in the table given in the ANSWER BOOK. (4)

2.2 How many people participated in this survey? (1)

2.3 Write down the modal class. (1)

2.4 Draw an ogive (cumulative frequency graph) to represent the data on the grid given in the ANSWER BOOK. (3)

2.5 Determine the percentage of marketing calls received by people older than 54 years. (3)

AGE OF PERSON IN SURVEY	FREQUENCY	CUMULATIVE FREQUENCY
$20 < x \leq 30$	7	7
$30 < x \leq 40$		27
$40 < x \leq 50$	25	
$50 < x \leq 60$		64
$60 < x \leq 70$		72
$70 < x \leq 80$	4	
$80 < x \leq 90$		80

THANK YOU

BELGIUM CAMPUS iTversity

It's the way we're wired

/belgiumcampusSA

#Belgium Campus

/belgiumcampus

info@belgiumcampus.ac.za +27 10 593 53 68

www.belgiumcampus.ac.za