CHAPTER 10

Statistics

- This chapter covers revision of central tendency in ungrouped data and then extends this to measures of central tendency in grouped data. The range is revised and extended to include percentiles, quartiles, interquartile and semi interquartile range. The five number summary and box and whisker diagram is introduced here. Finally statistical summaries are applied to data to make meaningful comments on the context associated with the data.
- Intervals for grouped data should be given using inequalities ($0 \leq x < 20$) rather than 0 - 19.
- Discuss the misuse of statistics in the real world and encourage awareness.

You can find data sets and statistics relevant to South Africa from the statssa website.

10.1 Collecting data

Exercise 10 – 1:

1. The following data set of dreams that learners have was collected from Grade 12 learners just after their final exams.
   {"I want to build a bridge!"; "I want to help the sick."; "I want running water!"}
   Categorise the data set.
   Solution: This data set cannot be written as numbers and so must be qualitative. This data set is anecdotal since it takes the form of a story. Therefore the data set is qualitative anecdotal.
2. The following data set of sweets in a packet was collected from visitors to a sweet shop.
   $\{23; 25; 22; 26; 27; 25; 21; 28\}$
   Categorise the data set.
   Solution: This data set is a set of numbers and so must be quantitative. This data set is discrete since it can be represented by integers and is a count of the number of sweets. Therefore the data set is quantitative discrete.
3. The following data set of questions answered correctly was collected from a class of maths learners.
   $\{3; 5; 2; 6; 7; 5; 1; 2\}$
   Categorise the data set.
   Solution: This data set is a set of numbers and so must be quantitative. This data set is discrete since it can be represented by integers and is a count of the number of questions answered correctly. Therefore the data set is quantitative discrete.

For more exercises, visit www.everythingmaths.co.za and click on 'Practise Maths'. 1. 2GM2 2. 2GM3 3. 2GM4

www.everythingmaths.co.za **m.everythingmaths.co.za**

10.2 Measures of central tendency

Mean

Median

Mode

Exercise 10 – 2:

1. Calculate the **mean** of the following data set:
   $\{9; 14; 9; 14; 8; 8; 9; 8; 9; 9\}$. Round your answer to 1 decimal place.
   Solution:

$$\begin{aligned} \text{mean} &= \frac{9+14+9+14+8+8+9+8+9+9}{10} \\ &= 9{,}7 \end{aligned}$$

The mean is: 9,7.

2. Calculate the **median** of the following data set:
   $\{4; 13; 10; 13; 13; 4; 2; 13; 13; 13\}$.
   Solution:
   We first need to order the data set:
   $\{2; 4; 4; 10; 13; 13; 13; 13; 13; 13\}$.
   Since there are an even number of values in this data set (10) the median lies between the fifth and sixth place:

$$\begin{aligned} \text{median} &= \frac{13+13}{2} \\ &= 13 \end{aligned}$$

   The median is: 13.

3. Calculate the **mode** of the following data set:
   $\{6; 10; 6; 6; 13; 12; 12; 7; 13; 6\}$
   Solution:
   We first sort the data set: $\{6; 6; 6; 6; 7; 10; 12; 12; 13; 13\}$. The mode is the value that occurs most often in the data set.
   Therefore the mode is: 6

4. Calculate the mean, median and mode of the following data sets:
   a) $\{2; 5; 8; 8; 11; 13; 22; 23; 27\}$
      Solution:
      The data set is already ordered.

$$\begin{aligned} \text{mean} &= \frac{2+5+8+8+11+13+22+23+27}{9} \\ &= 13{,}2 \end{aligned}$$

      Since there is an odd number of values in this data set the median lies at the fifth number: 11
      The mode is the value that occurs the most. In this data set the mode is 8.
      The mean, median and mode are: mean: 13,2; median: 11; mode: 8.
   b) $\{15; 17; 24; 24; 26; 28; 31; 43\}$
      Solution:
      The data set is already ordered.

$$\begin{aligned} \text{mean} &= \frac{15+17+24+24+26+28+31+43}{8} \\ &= 26 \end{aligned}$$

      Since there is an even number of values in this data set the median lies between the fourth and fifth numbers:

$$\begin{aligned} \text{median} &= \frac{24+26}{2} \\ &= 25 \end{aligned}$$

      The mode is the value that occurs the most. In this data set the mode is 24.
      The mean, median and mode are: mean: 26; median: 25; mode: 24.
   c) $\{4; 11; 3; 15; 11; 13; 25; 17; 2; 11\}$
      Solution:
      We first need to order the data set: $\{2; 3; 4; 11; 11; 11; 13; 15; 17; 25\}$.

$$\begin{aligned} \text{mean} &= \frac{2+3+4+11+11+11+13+15+17+25}{10} \\ &= 11{,}2 \end{aligned}$$

Since there is an even number of values in this data set the median lies between the fifth and sixth numbers:

$$\begin{aligned}\text{median} &= \frac{11+11}{2}\\ &= 11\end{aligned}$$

The mode is the value that occurs the most. In this data set the mode is 11.
Therefore the mean, median and mode are: mean: 11,2; median: 11; mode: 11.

d) $\{24; 35; 28; 41; 31; 49; 31\}$
Solution:
We first need to order the data set: $\{24; 28; 31; 31; 35; 41; 49\}$

$$\begin{aligned}\text{mean} &= \frac{24+28+31+31+35+41+49}{7}\\ &= 34{,}3\end{aligned}$$

Since there is an odd number of values in this data set the median lies at the fourth number: 31
The mode is the value that occurs the most. In this data set the mode is 31.
The mean, median and mode are: mean: 34,29; median: 31; mode: none.

5. The ages of 15 runners of the Comrades Marathon were recorded:

$$\{31; 42; 28; 38; 45; 51; 33; 29; 42; 26; 34; 56; 33; 46; 41\}$$

Calculate the mean, median and modal age.
Solution:
We first need to order the data set: $\{26; 28; 29; 31; 33; 33; 34; 38; 41; 42; 42; 45; 46; 51; 56\}$

$$\begin{aligned}\text{mean} &= \frac{26+28+29+31+33+33+34+38+41+42+42+45+46+51+56}{15}\\ &= 38{,}3\end{aligned}$$

Since there is an odd number of values in this data set the median lies at the eighth number: 38.
The mode is the value that occurs the most. In this data set there are two modes: 33 and 42.
Therefore the mean, median and modal ages are: mean: 38,3; median 38; mode 33 and 42.

6. A group of 10 friends each have some stones. They work out that the **mean** number of stones they have is 6. Then 7 friends leave with an unknown number (x) of stones. The remaining 3 friends work out that the **mean** number of stones they have left is 12,33.
When the 7 friends left, how many stones did they take with them?
Solution:
If the **mean** number of stones the group originally had was **6** then the total number of stones must have been:

$$\begin{aligned}\text{mean} &= \frac{\text{number of stones (before)}}{\text{group size}}\\ \text{number of stones (before)} &= \text{mean} \times \text{group size}\\ \text{number of stones (before)} &= (6) \times (10)\\ \text{number of stones (before)} &= 60\end{aligned}$$

We are then told that **7** friends leave and thereafter the **mean** number of stones left is **12,33**. Now we can work out the remaining number of stones.

$$\begin{aligned}\text{mean} &= \frac{\text{number of stones (after)}}{\text{group size}}\\ \text{number of stones (after)} &= \text{mean} \times \text{group size}\\ \text{number of stones (after)} &= (12{,}33) \times (3)\\ \text{number of stones (after)} &= 37\end{aligned}$$

Now we can calculate how many stones were taken by the 7 friends who left the group.

$$\text{number of stones removed } (x) = \text{items before} - \text{items after}$$
$$\text{number of stones removed } (x) = (60) - (37)$$
$$\text{number of stones removed } (x) = 23$$

7. A group of 9 friends each have some coins. They work out that the **mean** number of coins they have is 4. Then 5 friends leave with an unknown number (x) of coins. The remaining 4 friends work out that the **mean** number of coins they have left is 2,5.

   When the 5 friends left, how many coins did they take with them?

   Solution:

   If the **mean** number of coins the group originally had was **4** then the total number of coins must have been:

$$\text{mean} = \frac{\text{number of coins (before)}}{\text{group size}}$$
$$\text{number of coins (before)} = \text{mean} \times \text{group size}$$
$$\text{number of coins (before)} = (4) \times (9)$$
$$\text{number of coins (before)} = 36$$

   We are then told that **5** friends leave and thereafter the **mean** number of coins left is **2,5**. Let us work out the remaining number of coins.

$$\text{mean} = \frac{\text{number of coins (after)}}{\text{group size}}$$
$$\text{number of coins (after)} = \text{mean} \times \text{group size}$$
$$\text{number of coins (after)} = (2{,}5) \times (4)$$
$$\text{number of coins (after)} = 10$$

   Now we can calculate how many coins were taken by the 5 friends who left the group.

$$\text{number of coins removed } (x) = \text{items before} - \text{items after}$$
$$\text{number of coins removed } (x) = (36) - (10)$$
$$\text{number of coins removed } (x) = 26$$

8. A group of 9 friends each have some marbles. They work out that the **mean** number of marbles they have is 3. Then 3 friends leave with an unknown number (x) of marbles. The remaining 6 friends work out that the **mean** number of marbles they have left is 1,17.

   When the 3 friends left, how many marbles did they take with them?

   Solution:

   If the **mean** number of marbles the group originally had was **3** then the total number of marbles must have been:

$$\text{mean} = \frac{\text{number of marbles (before)}}{\text{group size}}$$
$$\text{number of marbles (before)} = \text{mean} \times \text{group size}$$
$$\text{number of marbles (before)} = (3) \times (9)$$
$$\text{number of marbles (before)} = 27$$

   We are then told that **3** friends leave and thereafter the **mean** number of marbles left is **1,17**. Let us work out the remaining number of marbles.

$$\text{mean} = \frac{\text{number of marbles (after)}}{\text{group size}}$$
$$\text{number of marbles (after)} = \text{mean} \times \text{group size}$$
$$\text{number of marbles (after)} = (1{,}17) \times (6)$$
$$\text{number of marbles (after)} = 7$$

   Now we can calculate how many marbles were taken by the 3 friends who left the group.

$$\text{number of marbles removed } (x) = \text{items before} - \text{items after}$$
$$\text{number of marbles removed } (x) = (27) - (7)$$
$$\text{number of marbles removed } (x) = 20$$

9. In the first of a series of jars, there is 1 sweet. In the second jar, there are 3 sweets. The mean number of sweets in the first two jars is 2.

   a) If the mean number of sweets in the first three jars is 3, how many sweets are there in the third jar?
   Solution:
   Let n_3 be the number of sweets in the third jar:

$$\begin{aligned} \frac{1+3+n_3}{3} &= 3 \\ 1+3+n_3 &= 9 \\ n_3 &= 5 \end{aligned}$$

   b) If the mean number of sweets in the first four jars is 4, how many sweets are there in the fourth jar?
   Solution:
   Let n_4 be the number of sweets in the fourth jar:

$$\begin{aligned} \frac{1+3+5+n_4}{4} &= 4 \\ 9+n_4 &= 16 \\ n_4 &= 7 \end{aligned}$$

10. Find a set of five ages for which the mean age is 5, the modal age is 2 and the median age is 3 years.
Solution:
Let the five different ages be $x_1,\ x_2,\ x_3,\ x_4$ and x_5. Therefore the mean is:

$$\begin{aligned} \frac{x_1+x_2+x_3+x_4+x_5}{5} &= 5 \\ x_1+x_2+x_3+x_4+x_5 &= 25 \end{aligned}$$

The median value is at position 3, therefore $x_3 = 3$.
The mode is the age that occurs most often. We have 5 ages to work with and we know one of the ages is 3 (from the median). So the ordered data set is: $\{x_1; x_2; 3; x_4; x_5\}$ (remember that we always calculate mean, mode and median using the ordered data set). We are told that the mode is 2. Looking at the ordered data set we see that either x_1 or x_2 must be 2 (x_4 and x_5 cannot be 2 as that would make the data set unordered). However, if only one of these values is 2 then the mode will not be 2. Therefore $x_1 = x_2 = 2$.
So we can now update our calculation of the mean:

$$\begin{aligned} 2+2+3+x_4+x_5 &= 25 \\ 18 &= x_4+x_5 \end{aligned}$$

x_4 and x_5 can be any numbers that add up to 18 and are not the same (if they were the same then the mode would not be 2), so 12 and 6 or 8 and 10 or 3 and 15, etc.
Possible data sets:

Data set 1: $\{2; 2; 3; 4; 14\}$
Data set 2: $\{2; 2; 3; 5; 13\}$
Data set 3: $\{2; 2; 3; 6; 12\}$
Data set 4: $\{2; 2; 3; 7; 11\}$
Data set 5: $\{2; 2; 3; 8; 10\}$

Note that the set of ages must be ordered, the median value must be 3 and there must be 2 ages of 2.

11. Four friends each have some marbles. They work out that the mean number of marbles they have is 10. One friend leaves with 4 marbles. How many marbles do the remaining friends have together?

Solution:

Let the number of marbles per friend be x_1, x_2, x_3 and x_4.

$$\frac{x_1 + x_2 + x_3 + x_4}{4} = 10$$
$$x_1 + x_2 + x_3 + x_4 = 40$$

One friend leaves:

$$x_1 + x_2 + x_3 = 40 - 4$$
$$x_1 + x_2 + x_3 = 36$$

Therefore the remaining friends have 36 marbles.

For more exercises, visit www.everythingmaths.co.za and click on 'Practise Maths'.

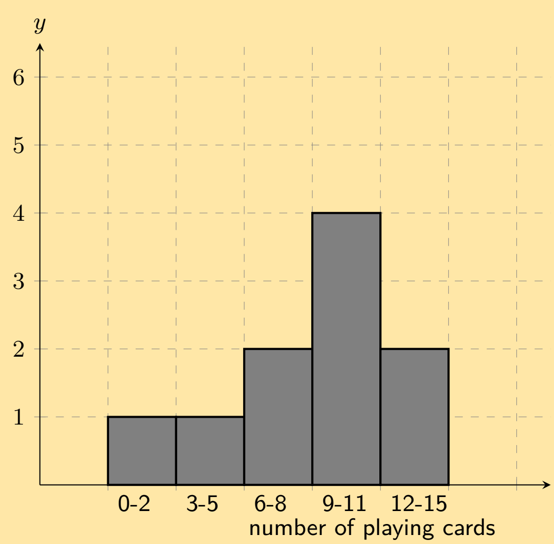
1. 2GM6 2. 2GM7 3. 2GM8 4a. 2GM9 4b. 2GMB 4c. 2GMC
4d. 2GMD 5. 2GMF 6. 2GMG 7. 2GMH 8. 2GMJ 9. 2GMK
10. 2GMM 11. 2GMN

www.everythingmaths.co.za m.everythingmaths.co.za

10.3 Grouping data

Exercise 10 – 3:

1. A group of 10 learners count the number of playing cards they each have. This is a histogram describing the data they collected:

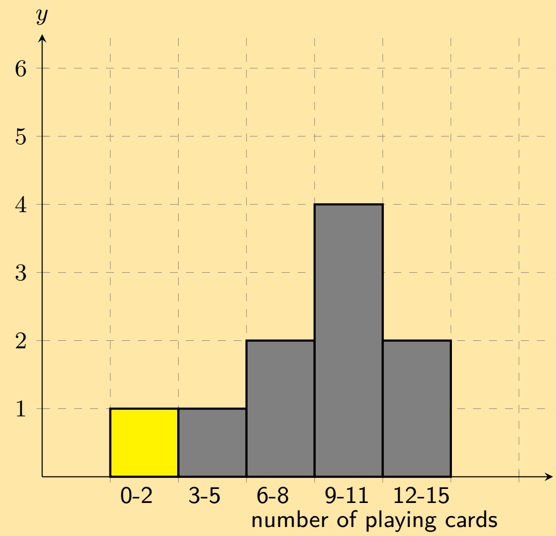


Count the number of playing cards in the following range: $0 \leq$ number of playing cards ≤ 2

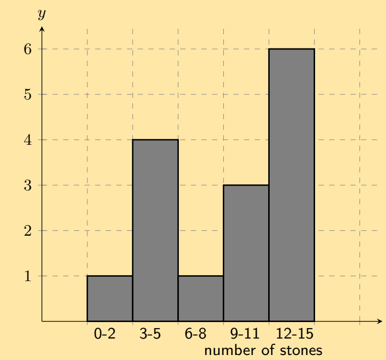
Solution:

From the graph the answer is: 1

From the histogram, we arrive at our answer by reading the height of the specified interval from the histogram.



2. A group of 15 learners count the number of stones they each have. This is a histogram describing the data they collected:

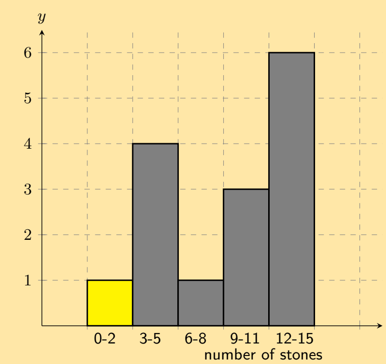


Count the number of stones in the following range: $0 \leq$ number of stones ≤ 2

Solution:

From the graph the answer is: 1

From the histogram, we arrive at our answer by reading the height of the specified interval from the histogram.



3. A group of **20** learners count the number of playing cards they each have. This is the data they collect:

14	9	11	8	13
2	3	4	16	17
9	19	10	14	4
16	16	11	2	17

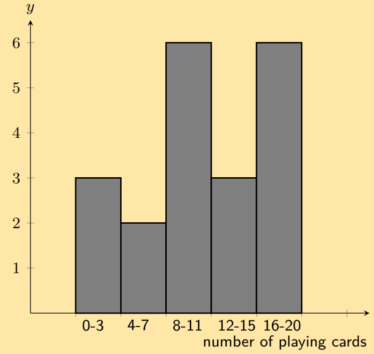
Count the number of learners who have from 12 up to 15 playing cards. In other words, how many learners have playing cards in the following range: $12 \leq$ number of playing cards ≤ 15? It may be helpful for you to draw a histogram in order to answer the question.

Solution:

Firstly we sort the table into sequential order, starting with the smallest value.

2	2	3	4	4
8	9	9	10	11
11	13	14	14	16
16	16	17	17	19

Secondly, we draw a histogram of the data:



From the histogram you can see that the number of learners with playing cards in the range: $12 \leq$ number of playing cards ≤ 15 is 3.

4. A group of **20** learners count the number of stones they each have. This is the data they collect:

16	6	11	19	20
17	13	1	5	12
5	2	16	11	16
6	10	13	6	17

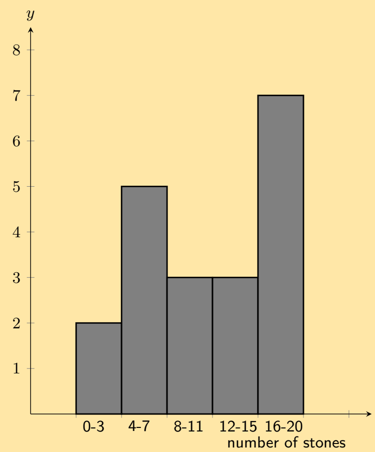
Count the number of learners who have from 4 up to 7 stones. In other words, how many learners have stones in the following range: $4 \leq$ number of stones ≤ 7? It may be helpful for you to draw a histogram in order to answer the question.

Solution:

Firstly we sort the table into sequential order, starting with the smallest value.

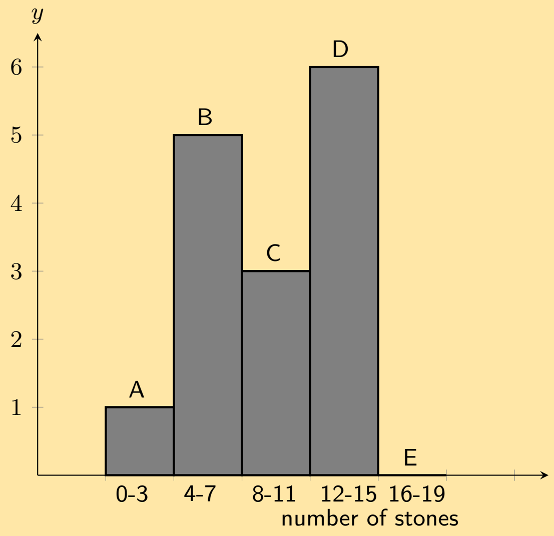
1	2	5	5	6
6	6	10	11	11
12	13	13	16	16
16	17	17	19	20

Secondly, we draw a histogram of the data:



From the histogram you can see that the number of learners with stones in the range: $4 \leq$ number of stones ≤ 7 is 5.

5. A group of 20 learners count the number of stones they each have. The learners draw a histogram describing the data they collected. However, they have made a mistake in drawing the histogram.



The data set below shows the correct information for the number of stones the learners have. Each value represents the number of stones for one learner.

$$\{4; 12; 15; 14; 18; 12; 17; 15; 1; 6; 6; 12; 6; 8; 6; 8; 17; 19; 16; 8\}$$

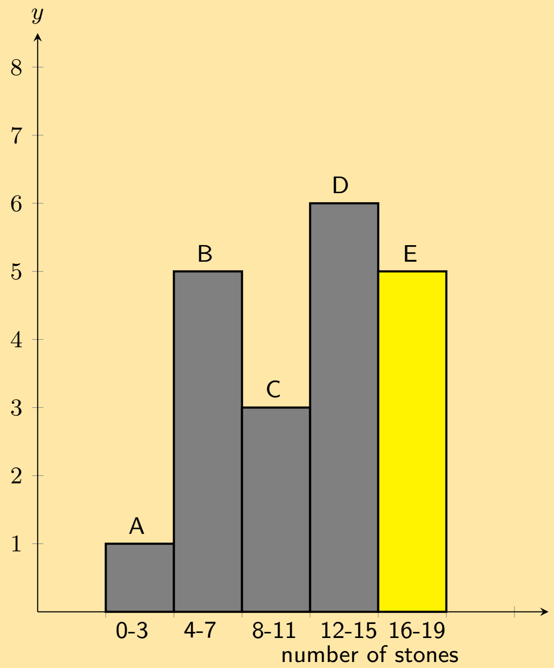
Help them figure out which **column** in the histogram is **incorrect**.

Solution:

We first need to order the data:

$$\{1; 4; 6; 6; 6; 6; 8; 8; 8; 12; 12; 12; 14; 15; 15; 16; 17; 17; 18; 19\}$$

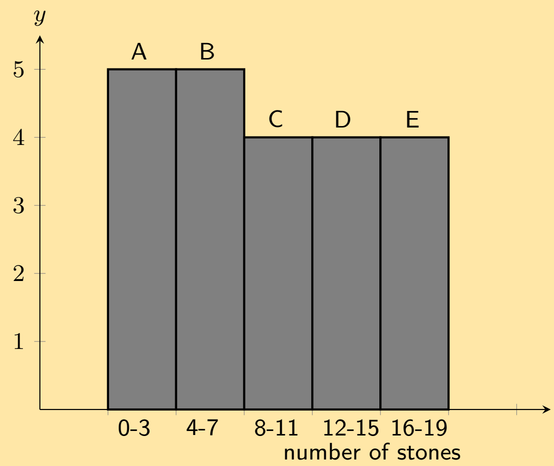
Using the ordered data set we can group the data and draw the correct histogram:



The column with the error in it was: E.

The learners used the incorrect value of **0**, when the correct value is 5.

6. A group of 20 learners count the number of stones they each have. The learners draw a histogram describing the data they collected. However, they have made a mistake in drawing the histogram.



The data set below shows the correct information for the number of stones the learners have. Each value represents the number of stones for one learner.

$$\{19; 11; 5; 2; 3; 4; 14; 2; 12; 19; 11; 14; 2; 19; 11; 5; 17; 10; 1; 12\}$$

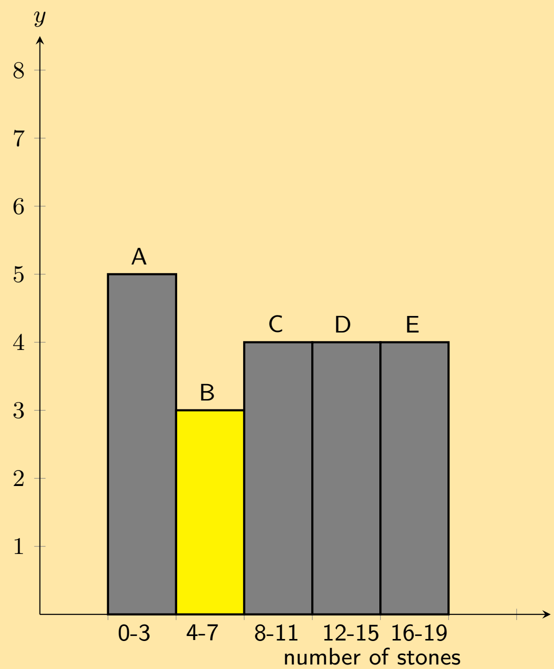
Help them figure out which **column** in the histogram is **incorrect**.

Solution:

We first need to order the data:

$$\{1; 2; 2; 2; 3; 4; 5; 5; 10; 11; 11; 11; 12; 12; 14; 14; 17; 19; 19; 19\}$$

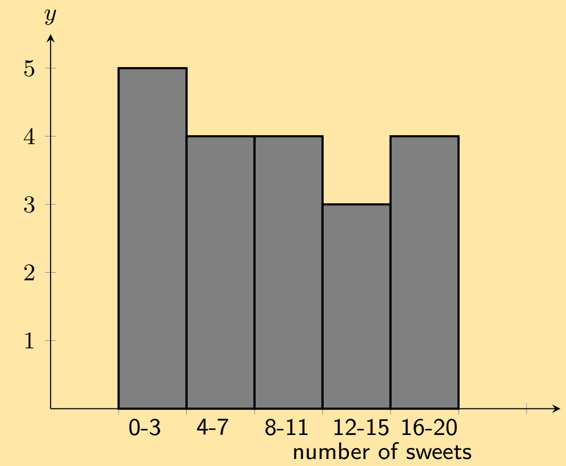
Using the ordered data set we can group the data and draw the correct histogram:



The column with the error in it was: B.

The learners used the incorrect value of **5**, when the correct value is 3.

7. A group of learners count the number of sweets they each have. This is a histogram describing the data they collected:



A cat jumps onto the table, and all their notes land on the floor, mixed up, by accident!
Help them find which of the following data sets match the above histogram:
Data Set A

2	1	20	10	5
3	10	2	6	1
2	2	17	3	18
3	7	10	8	18

Data Set B

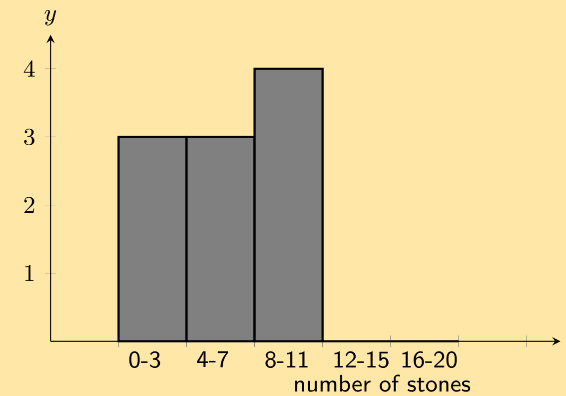
2	9	12	10	5
9	9	10	13	6
5	11	10	7	7

Data Set C

3	12	16	10	15
17	18	2	3	7
11	12	8	2	7
17	3	11	4	4

Solution:
The correct answer is: Data Set C

8. A group of learners count the number of stones they each have. This is a histogram describing the data they collected:



A cleaner knocks over their table, and all their notes land on the floor, mixed up, by accident!
Help them find which of the following data sets match the above histogram:
Data Set A

12	4	2	15	10
18	10	16	16	19
1	2	9	10	16
10	11	9	2	13

Data Set B

7	10	4	5	8
7	12	10	14	5
1	9	2	13	3

Data Set C

9	3	8	5	8
5	8	1	4	3

Solution:

The correct answer is: Data Set C

9. A class experiment was conducted and 50 learners were asked to guess the number of sweets in a jar. The following guesses were recorded:

56	49	40	11	33	33	37	29	30	59
21	16	38	44	38	52	22	24	30	34
42	15	48	33	51	44	33	17	19	44
47	23	27	47	13	25	53	57	28	23
36	35	40	23	45	39	32	58	22	40

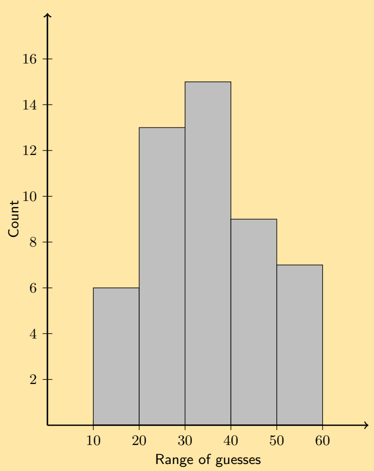
a) Draw up a grouped frequency table using the intervals $10 < x \leq 20$, $20 < x \leq 30$, $30 < x \leq 40$, $40 < x \leq 50$ and $50 < x \leq 60$.

Solution:

Group	Frequency
$10 < x \leq 20$	6
$20 < x \leq 30$	13
$30 < x \leq 40$	15
$40 < x \leq 50$	9
$50 < x \leq 60$	7

b) Draw the histogram corresponding to the frequency table of the grouped data.

Solution:



For more exercises, visit www.everythingmaths.co.za and click on 'Practise Maths'.

1. 2GMQ 2. 2GMR 3. 2GMS 4. 2GMT 5. 2GMV 6. 2GMW
7. 2GMX 8. 2GMY 9. 2GMZ

www.everythingmaths.co.za **m.everythingmaths.co.za**

Exercise 10 – 4:

1. Consider the following grouped data and calculate the mean, the modal group and the median group.

Mass (kg)	**Count**
$40 < m \leq 45$	7
$45 < m \leq 50$	10
$50 < m \leq 55$	15
$55 < m \leq 60$	12
$60 < m \leq 65$	6

Solution:
To find the mean we use the middle value for each group. The count then tells us how many times that value occurs in the data set. Therefore the mean is:

$$\begin{aligned}\text{mean} &= \frac{7(43)+10(48)+15(53)+12(58)+6(63)}{7+10+15+12+6}\\ &= \frac{2650}{50}\\ &= 53\end{aligned}$$

The modal group is the group with the highest number of data values. This is $50 < m \leq 55$ with 15 data values.
The median group is the central group. There are 5 groups and so the central group is the third one: $50 < m \leq 55$.
Mean: 52; Modal group: $50 < m \leq 55$; Median group: $50 < m \leq 55$.

2. Find the mean, the modal group and the median group in this data set of how much time people needed to complete a game.

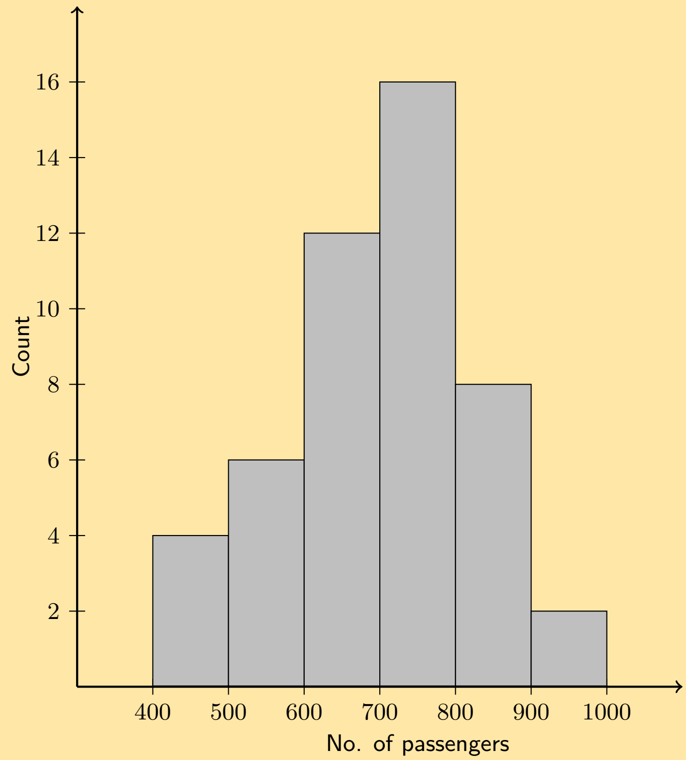
Time (s)	**Count**
$35 < t \leq 45$	5
$45 < t \leq 55$	11
$55 < t \leq 65$	15
$65 < t \leq 75$	26
$75 < t \leq 85$	19
$85 < t \leq 95$	13
$95 < t \leq 105$	6

Solution:
To find the mean we use the middle value for each group. The count then tells us how many times that value occurs in the data set. Therefore the mean is:

$$\begin{aligned}\text{mean} &= \frac{5(40{,}5)+11(50{,}5)+15(60{,}5)+26(70{,}5)+19(80{,}5)+13(90{,}5)+6(100{,}5)}{5+11+15+26+19+13+6}\\ &= \frac{6807{,}5}{95}\\ &= 71{,}66\end{aligned}$$

The modal group is the group with the highest number of data values. This is $65 < m \leq 75$ with 26 data values.
The median group is the central group. There are 7 groups and so the central group is the fourth one: $65 < m \leq 75$.
Mean: 70,66; Modal group: $65 < t \leq 75$; Median group: $65 < t \leq 75$.

3. The histogram below shows the number of passengers that travel in Alfred's minibus taxi per week.



Calculate:

a) the modal interval

Solution:

The modal interval is the interval with the highest number of data values. For this data set it is: $700 < x \leq 800$ with 16 values.

b) the total number of passengers to travel in Alfred's taxi

Solution:

We add up the counts in each group and then multiply these counts with the central value for each group: $4(450) + 6(550) + 12(650) + 16(750) + 8(850) + 2(950) = 33\ 600$.

c) an estimate of the mean

Solution:

There are 48 values in the data set. Therefore the mean is $\frac{33\ 600}{48} = 700$.

d) an estimate of the median

Solution:

We are looking for an estimate of the median rather than the median group here. In this case we note that there are 48 data values in the data set. Therefore the median will lie between the 24^{th} and 25^{th} values.

We note that 22 values in the first 3 groups and 38 values in the first four groups so the median must lie in the fourth group. Therefore we can estimate the median as the middle value of the fourth group: 750.

e) if it is estimated that every passenger travelled an average distance of 5 km, how much money would Alfred have made if he charged R 3,50 per km?

Solution:

$3{,}50 \times 5 \times 33\ 600 =$ R 588 000.

For more exercises, visit www.everythingmaths.co.za and click on 'Practise Maths'. 1. 2GN2 2. 2GN3 3. 2GN4

www.everythingmaths.co.za m.everythingmaths.co.za

10.4 Measures of dispersion

Range

Percentiles

Percentiles for grouped data

Ranges

Exercise 10 – 5:

1. A group of **15** learners count the number of sweets they each have. This is the data they collect:

$$\begin{array}{ccccc} 4 & 11 & 14 & 7 & 14 \\ 5 & 8 & 7 & 12 & 12 \\ 5 & 13 & 10 & 6 & 7 \end{array}$$

   Calculate the **range** of values in the data set.
   Solution:
   We first need to order the data set:

$$\{4; 5; 5; 6; 7; 7; 7; 8; 10; 11; 12; 12; 13; 14; 14\}$$

   Next we find the maximum value in the data set:

$$\text{maximum value } = 14$$

   Then we find the minimum value in the data set:

$$\text{minimum value } = 4$$

   Finally, we calculate the range of the data set:

$$\begin{aligned} \text{range } &= (\text{maximum value}) - (\text{minimum value}) \\ &= (14) - (4) \\ &= 10 \end{aligned}$$

2. A group of **10** learners count the number of playing cards they each have. This is the data they collect:

$$\begin{array}{ccccc} 5 & 1 & 3 & 1 & 4 \\ 10 & 1 & 3 & 3 & 4 \end{array}$$

   Calculate the **range** of values in the data set.
   Solution:
   We first need to order the data set:

$$\{1; 1; 1; 3; 3; 3; 4; 4; 5; 10\}$$

   Next we find the maximum value in the data set:

$$\text{maximum value } = 10$$

   Then we find the minimum value in the data set:

$$\text{minimum value } = 1$$

Finally, we calculate the range of the data set:

$$\begin{aligned}\text{range} &= (\text{maximum value}) - (\text{minimum value})\\ &= 10 - 1\\ &= 9\end{aligned}$$

3. Find the range of the data set

$$\{1; 2; 3; 4; 4; 4; 5; 6; 7; 8; 8; 9; 10; 10\}$$

Solution:
The data set is already ordered.
Firstly, we find the maximum value in the data set:

$$\text{maximum value } = 10$$

Secondly, we find the minimum value in the data set:

$$\text{minimum value } = 1$$

Finally, we calculate the range of the data set:

$$\begin{aligned}\text{range} &= (\text{maximum value}) - (\text{minimum value})\\ &= 10 - 1\\ &= 9\end{aligned}$$

4. What are the quartiles of this data set?

$$\{3; 5; 1; 8; 9; 12; 25; 28; 24; 30; 41; 50\}$$

Solution:
We first order the data set.

$$\{1; 3; 5; 8; 9; 12; 24; 25; 28; 30; 41; 50\}$$

Next we find the ranks of the quartiles. Using the percentile formula with $n = 12$, we can find the rank of the 25^{th}, 50^{th} and 75^{th} percentiles:

$$\begin{aligned}r_{25} &= \frac{25}{100}(12-1)+1\\ &= 3{,}75\\ r_{50} &= \frac{50}{100}(12-1)+1\\ &= 6{,}5\\ r_{75} &= \frac{75}{100}(12-1)+1\\ &= 9{,}25\end{aligned}$$

Find the values of the quartiles. Note that each of these ranks is a fraction, meaning that the value for each percentile is somewhere in between two values from the data set.
For the 25^{th} percentile the rank is 3,75, which is between the third and fourth values. Therefore the 25^{th} percentile is $\frac{5+8}{2} = 6{,}5$.
For the 50^{th} percentile (the median) the rank is 6,5, meaning halfway between the sixth and seventh values. Therefore the median is $\frac{12+24}{2} = 18$. For the 75^{th} percentile the rank is 9,25, meaning between the ninth and tenth values. Therefore the 75^{th} percentile is $\frac{28+30}{2} = 29$.
Therefore we get the following values for the quartiles: $Q_1 = 6{,}5$; $Q_2 = 18$; $Q_3 = 29$.

5. A class of 12 learners writes a test and the results are as follows:

$$\{20; 39; 40; 43; 43; 46; 53; 58; 63; 70; 75; 91\}$$

Find the range, quartiles and the interquartile range.

Solution:

The data set is ordered.

The range is:

$$\begin{aligned} \text{range} &= (\text{maximum value}) - (\text{minimum value}) \\ &= (91) - (20) \\ &= 71 \end{aligned}$$

To find the quartiles we start by finding the ranks of the quartiles. Using the percentile formula with $n = 12$, we can find the rank of the 25^{th}, 50^{th} and 75^{th} percentiles:

$$\begin{aligned} r_{25} &= \frac{25}{100}(12-1)+1 \\ &= 3{,}75 \\ r_{50} &= \frac{50}{100}(12-1)+1 \\ &= 6{,}5 \\ r_{75} &= \frac{75}{100}(12-1)+1 \\ &= 9{,}25 \end{aligned}$$

Find the values of the quartiles. Note that each of these ranks is a fraction, meaning that the value for each percentile is somewhere in between two values from the data set.

For the 25^{th} percentile the rank is 3,75, which is between the third and fourth values. Therefore the 25^{th} percentile is $\frac{40+43}{2} = 41{,}5$.

For the 50^{th} percentile (the median) the rank is 6,5, meaning halfway between the sixth and seventh values. Therefore the median is $\frac{46+53}{2} = 49{,}5$. For the 75^{th} percentile the rank is 9,25, meaning between the ninth and tenth values. Therefore the 75^{th} percentile is $\frac{63+70}{2} = 66{,}5$.

Therefore we get the following values for the quartiles: $Q_1 = 41{,}5$; $Q_2 = 49{,}5$; $Q_3 = 66{,}5$.

Interquartile range:

$$\begin{aligned} \text{interquartile range} &= \text{quartile 3} - \text{quartile 1} \\ &= 66{,}5 - 41{,}5 \\ &= 25 \end{aligned}$$

6. Three sets of data are given:

Data set 1: $\{9; 12; 12; 14; 16; 22; 24\}$

Data set 2: $\{7; 7; 8; 11; 13; 15; 16\}$

Data set 3: $\{11; 15; 16; 17; 19; 22; 24\}$

For each data set find:

a) the range

Solution:

All three data sets are ordered. To find the range we subtract the minimum value from the maximum value. Doing so for each data set gives the following values for the range.

Data set 1: $24 - 9 = 15$

Data set 2: $16 - 7 = 9$

Data set 3: $24 - 11 = 13$

b) the lower quartile

Solution:

For each data set $n = 7$. Therefore the rank of the 25^{th} percentile is the same for each data set: $r_{25} = \frac{25}{100}(7-1)+1 = 2{,}5$. Therefore for each data set the lower quartile lies between the second and third values.

The lower quartile for each data set is:
Data set 1: 12
Data set 2: 7,5
Data set 3: 15,5

c) the median

Solution:

For each data set $n = 7$. Therefore the rank of the 50^{th} percentile is the same for each data set: $r_{50} = \frac{50}{100}(7-1)+1 = 4$. Therefore for each data set the median is the fourth value.
The median for each data set is:
Data set 1: 14
Data set 2: 11
Data set 3: 17

d) the upper quartile

Solution:

For each data set $n = 7$. Therefore the rank of the 75^{th} percentile is the same for each data set: $r_{75} = \frac{75}{100}(7-1)+1 = 5{,}5$. Therefore for each data set the lower quartile lies between the fifth and sixth values.
The upper quartile for each data set is:
Data set 1: 19
Data set 2: 14
Data set 3: 20,5

e) the interquartile range

Solution:

The interquartile range is calculated by subtracting the lower quartile from the upper quartile.
Data set 1: $19 - 12 = 7$
Data set 2: $14 - 7{,}5 = 6{,}5$
Data set 3: $20{,}5 - 15{,}5 = 5$

f) the semi-interquartile range

Solution:

The semi-interquartile range is half the interquartile range.
Data set 1: $\frac{7}{2} = 3{,}5$
Data set 2: $\frac{6{,}5}{2} = 3{,}25$
Data set 3: $\frac{5}{2} = 2{,}5$

For more exercises, visit www.everythingmaths.co.za and click on 'Practise Maths'.
1. 2GN5 2. 2GN6 3. 2GN7 4. 2GN8 5. 2GN9 6. 2GNB

www.everythingmaths.co.za m.everythingmaths.co.za

10.5 Five number summary

Wikihow shows a summary with short animations of how to make a box and whisker plot.

Exercise 10 – 6:

1. Lisa is working in a computer store. She sells the following number of computers each month:

$$\{27; 39; 3; 15; 43; 27; 19; 54; 65; 23; 45; 16\}$$

Give the five number summary and box-and-whisker plot of Lisa's sales.

Solution:

We first order the data set.

$$\{3; 15; 16; 19; 23; 27; 27; 39; 43; 45; 54; 65\}$$

Now we can read off the minimum as the first value (3) and the maximum as the last value (65).

Next we need to determine the quartiles.

There are 12 values in the data set. Using the percentile formula, we can determine that the median lies between the sixth and seventh values, making it:

$$\frac{27+27}{2} = 27$$

The first quartile lies between the third and fourth values, making it:

$$\frac{16+19}{2} = 17{,}5$$

The third quartile lies between the ninth and tenth values, making it:

$$\frac{43+45}{2} = 44$$

This provides the five number summary of the data set and allows us to draw the following box-and-whisker plot.

Five number summary:

Minimum: 3

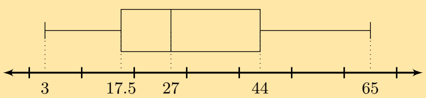
Q_1: 17,5

Median: 27

Q_3: 44

Maximum: 65

Box-and-whisker plot:



2. Zithulele works as a telesales person. He keeps a record of the number of sales he makes each month. The data below show how much he sells each month.

$$\{49; 12; 22; 35; 2; 45; 60; 48; 19; 1; 43; 12\}$$

Give the five number summary and box-and-whisker plot of Zithulele's sales.

Solution:

We first order the data set.

$$\{1; 2; 12; 12; 19; 22; 35; 43; 45; 48; 49; 60\}$$

Now we can read off the minimum as the first value (1) and the maximum as the last value (60).

Next we need to determine the quartiles.

There are 12 values in the data set. Using the percentile formula, we can determine that the median lies between the sixth and seventh values, making it:

$$\frac{22+35}{2} = 28{,}5$$

The first quartile lies between the third and fourth values, making it:

$$\frac{12+12}{2} = 12$$

The third quartile lies between the ninth and tenth values, making it:

$$\frac{45+48}{2} = 46{,}5$$

The five number summary is:

Minimum: 1

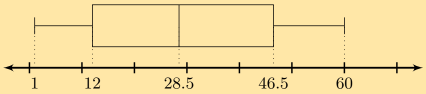
Q_1: 12

Median: 28,5

Q_3: 46,5

Maximum: 60

The box and whisker plot is:



3. Nombusa has worked as a florist for nine months. She sold the following number of wedding bouquets:

$$\{16; 14; 8; 12; 6; 5; 3; 5; 7\}$$

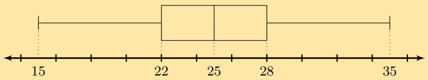
Give the five number summary of Nombusa's sales.
Solution:
We first order the data set.

$$\{3; 5; 5; 6; 7; 8; 12; 14; 16\}$$

Now we can read off the minimum as the first value (3) and the maximum as the last value (16).
Next we need to determine the quartiles.
There are 9 values in the data set. Using the percentile formula, we can determine that the median lies at the fifth value, making it 7.
The first quartile lies at the third value, making it 5.
The third quartile lies at the seventh values, making it 12.
The five number summary is:
Minimum: 3
Q_1: 5
Median: 7
Q_3: 12
Maximum: 16

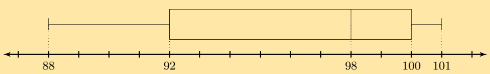
4. Determine the five number summary for each of the box-and-whisker plots below.

a)



Solution:
The box shows the interquartile range (the distance between $Q1$ and $Q3$). A line inside the box shows the median. The lines extending outside the box (the whiskers) show where the minimum and maximum values lie. Reading off the graph we obtain the following five number summary:
Minimum: 15
Q_1: 22
Median: 25
Q_3: 28
Maximum: 35

b)



Solution:
The box shows the interquartile range (the distance between $Q1$ and $Q3$). A line inside the box shows the median. The lines extending outside the box (the whiskers) show where the minimum and maximum values lie. Reading off the graph we obtain the following five number summary:
Minimum: 88
Q_1: 92
Median: 98
Q_3: 100
Maximum: 101

For more exercises, visit www.everythingmaths.co.za and click on 'Practise Maths'.
1. 2GND 2. 2GNF 3. 2GNG 4a. 2GNH 4b. 2GNJ

www.everythingmaths.co.za m.everythingmaths.co.za

End of chapter Exercise 10 – 7:

1. The following data set of heights was collected from a class of learners.
   {1,70 m; 1,41 m; 1,60 m; 1,32 m; 1,80 m; 1,40 m}
   Categorise the data set.
   Solution:
   This data set is a set of numbers and so must be quantitative.
   This data set is continuous since it cannot be represented by integers.
   Therefore the data set is quantitative continuous.
2. The following data set of sandwich spreads was collected from learners at lunch.
   {cheese; peanut butter; jam; cheese; honey}
   Categorise the data set.
   Solution:
   This data set cannot be written as numbers and so must be qualitative.
   This data set is categorical since it comes from a limited set of possibilities.
   Therefore the data set is qualitative categorical. para>
3. Calculate the **mode** of the following data set:
   $\{10; 10; 10; 18; 7; 10; 3; 10; 7; 10; 7\}$
   Solution:
   We first sort the data set: $\{3; 7; 7; 7; 10; 10; 10; 10; 10; 10; 18\}$. The mode is the value that occurs most often in the data set.
   Therefore the mode is: 10.
4. Calculate the **median** of the following data set:
   $\{5; 5; 10; 7; 10; 2; 16; 10; 10; 10; 7\}$
   Solution:
   We first need to order the data set:
   $\{2; 5; 5; 7; 7; 10; 10; 10; 10; 10; 16\}$.
   Since there are an odd number of values in this data set (11) the median lies at the sixth place.
   The median is: 10.
5. In a park, the tallest 7 trees have heights (in metres):

   $$\{41; 60; 47; 42; 44; 42; 47\}$$

   Find the median of their heights.
   Solution:
   We first need to order the data set:
   $\{41; 42; 42; 44; 47; 47; 60\}$.
   Since there are an odd number of values in this data set (7) the median lies at the fourth place.
   The median is: 44.
6. The learners in Ndeme's class have the following ages:

   $$\{5; 6; 7; 5; 4; 6; 6; 6; 7; 4\}$$

   Find the mode of their ages.
   Solution:
   We first sort the data set: $\{4; 4; 5; 5; 6; 6; 6; 6; 7; 7\}$. The mode is the value that occurs most often in the data set.
   Therefore the mode is: 6.
7. A group of 7 friends each have some sweets. They work out that the **mean** number of sweets they have is 6. Then 4 friends leave with an unknown number (x) of sweets. The remaining 3 friends work out that the **mean** number of sweets they have left is 10,67.
   When the 4 friends left, how many sweets did they take with them?
   Solution:
   If the **mean** number of sweets the group originally had was **6** then the total number of sweets must have been:

$$\begin{aligned}
\text{mean} &= \frac{\text{number of sweets (before)}}{\text{group size}} \\
\text{number of sweets (before)} &= \text{mean} \times \text{group size} \\
\text{number of sweets (before)} &= (6) \times (7) \\
\text{number of sweets (before)} &= 42
\end{aligned}$$

We are then told that **4** friends leave and thereafter the **mean** number of sweets left is **10,67**. Let us work out the remaining number of sweets.

$$\begin{aligned}
\text{mean} &= \frac{\text{number of sweets (after)}}{\text{group size}} \\
\text{number of sweets (after)} &= \text{mean} \times \text{group size} \\
\text{number of sweets (after)} &= (10{,}67) \times (3) \\
\text{number of sweets (after)} &= 32
\end{aligned}$$

Now we can calculate how many sweets were taken by the 4 friends who left the group.

$$\begin{aligned}
\text{number of sweets removed } (x) &= \text{items before} - \text{items after} \\
\text{number of sweets removed } (x) &= (42) - (32) \\
\text{number of sweets removed } (x) &= 10
\end{aligned}$$

8. A group of 10 friends each have some sweets. They work out that the **mean** number of sweets they have is 3. Then 5 friends leave with an unknown number (x) of sweets. The remaining 5 friends work out that the **mean** number of sweets they have left is 3.
When the 5 friends left, how many sweets did they take with them?
Solution:
If the **mean** number of sweets the group originally had was **3** then the total number of sweets must have been:

$$\begin{aligned}
\text{mean} &= \frac{\text{number of sweets (before)}}{\text{group size}} \\
\text{number of sweets (before)} &= \text{mean} \times \text{group size} \\
\text{number of sweets (before)} &= (3) \times (10) \\
\text{number of sweets (before)} &= 30
\end{aligned}$$

We are then told that **5** friends leave and thereafter the **mean** number of sweets left is **3**. Let us work out the remaining number of sweets.

$$\begin{aligned}
\text{mean} &= \frac{\text{number of sweets (after)}}{\text{group size}} \\
\text{number of sweets (after)} &= \text{mean} \times \text{group size} \\
\text{number of sweets (after)} &= (3) \times (5) \\
\text{number of sweets (after)} &= 15
\end{aligned}$$

Now we can calculate how many sweets were taken by the 5 friends who left the group.

$$\begin{aligned}
\text{number of sweets removed } (x) &= \text{items before} - \text{items after} \\
\text{number of sweets removed } (x) &= (30) - (15) \\
\text{number of sweets removed } (x) &= 15
\end{aligned}$$

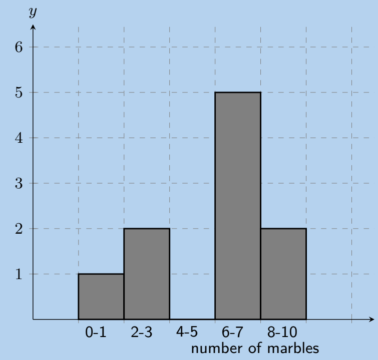
9. Five data values are represented as follows: $3x; x + 2; x{-}3; x + 4; 2x{-}5$, with a mean of 30. Solve for x.
Solution:

$$\bar{x} = \frac{x_1 + x_2 + ...x_n}{N}$$
$$30 = \frac{3x + x + 2 + x - 3 + x + 4 + 2x - 5}{5}$$
$$150 = 8x - 2$$
$$152 = 8x$$
$$x = \frac{152}{8}$$
$$\therefore x = 19$$

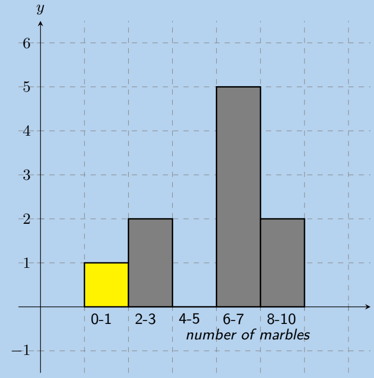
10. Five data values are represented as follows: $p + 1; p + 2; p + 9$. Find the mean in terms of p.
Solution:

$$\bar{x} = \frac{x_1 + x_2 + ...x_n}{N}$$
$$= \frac{p + 1 + p + 2 + p + 9}{3}$$
$$= \frac{3p + 12}{3}$$
$$\therefore \bar{x} = p + 4$$

11. A group of 10 learners count the number of marbles they each have. This is a histogram describing the data they collected:



Count the number of marbles in the following range: $0 \leq$ number of marbles ≤ 1
Solution:
From the graph the answer is: 1
From the histogram, we arrive at our answer by reading the height of the specified interval from the histogram.



12. A group of **20** learners count the number of playing cards they each have. This is the data they collect:

12	1	5	4	17	14	7	5	1	3
9	4	12	17	5	19	1	19	7	15

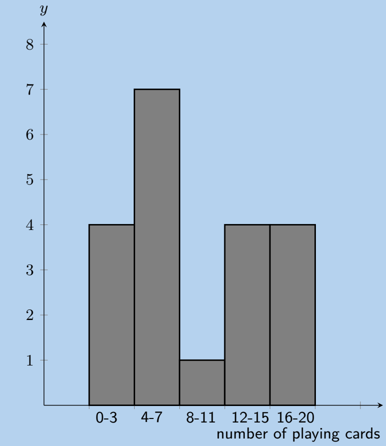
Count the number of learners who have from 0 up to 3 playing cards. In other words, how many learners have playing cards in the following range: $0 \leq$ number of playing cards ≤ 3? It may be helpful for you to draw a histogram in order to answer the question.

Solution:

Firstly we sort the table into sequential order, starting with the smallest value.

1	1	1	3	4	4	5	5	5	7
7	9	12	12	14	15	17	17	19	19

Secondly, we draw a histogram of the data:



From the histogram you can see that the number of learners with playing cards in the range: $0 \leq$ number of playing cards ≤ 3 is 4.

13. A group of **20** learners count the number of coins they each have. This is the data they collect:

17	11	1	15	14	3	4	18	5	14
18	19	4	18	15	16	13	20	8	18

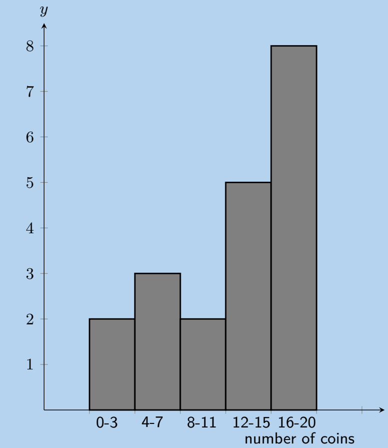
Count the number of learners who have from 4 up to 7 coins. In other words, how many learners have coins in the following range: $4 \leq$ number of coins ≤ 7? It may be helpful for you to draw a histogram in order to answer the question.

Solution:

Firstly we sort the table into sequential order, starting with the smallest value.

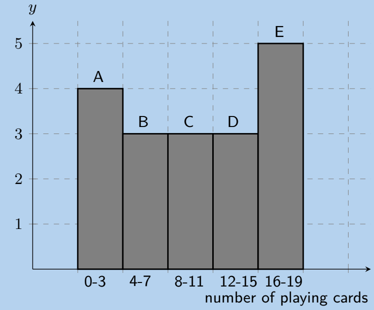
1	3	4	4	5	8	11	13	14	14
15	15	16	17	18	18	18	18	19	20

Secondly, we draw a histogram of the data:



From the histogram you can see that the number of learners with coins in the range: $4 \leq$ number of coins ≤ 7 is 3.

14. A group of 20 learners count the number of playing cards they each have. The learners draw a histogram describing the data they collected. However, they have made a mistake in drawing the histogram.



The data set below shows the correct information for the number of playing cards the learners have. Each value represents the number of playing cards for one learner.

$$\{18; 10; 3; 2; 19; 15; 2; 13; 11; 14; 10; 3; 5; 9; 4; 18; 11; 18; 16; 5\}$$

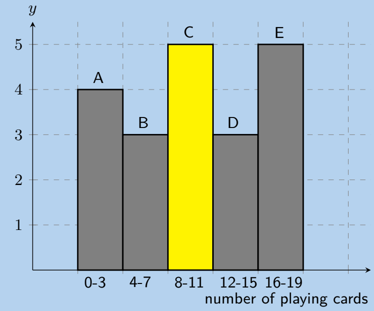
Help them figure out which **column** in the histogram is **incorrect**.

Solution:

We first need to order the data:

$$\{2; 2; 3; 3; 4; 5; 5; 9; 10; 10; 11; 11; 13; 14; 15; 16; 18; 18; 18; 19\}$$

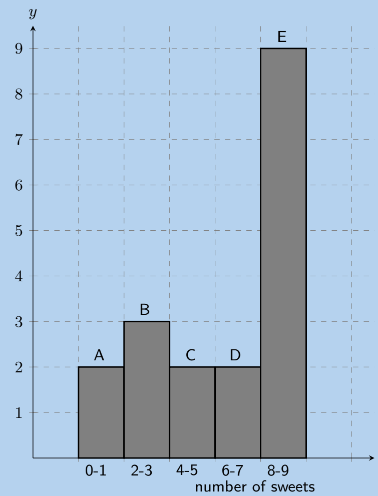
Using the ordered data set we can group the data and draw the correct histogram:



The column with the error in it was: C.

The learners used the incorrect value of **3**, when the correct value is 5.

15. A group of 10 learners count the number of sweets they each have. The learners draw a histogram describing the data they collected. However, they have made a mistake in drawing the histogram.



The data set below shows the correct information for the number of sweets the learners have. Each value represents the number of sweets for one learner.

$$\{1; 3; 7; 4; 5; 8; 2; 2; 1; 7\}$$

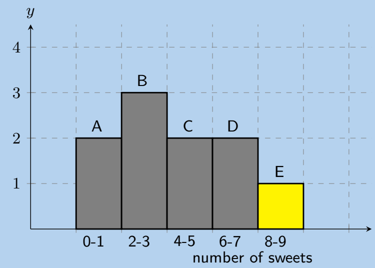
Help them figure out which **column** in the histogram is **incorrect**.

Solution:

We first need to order the data:

$$\{1; 1; 2; 2; 3; 4; 5; 7; 7; 8\}$$

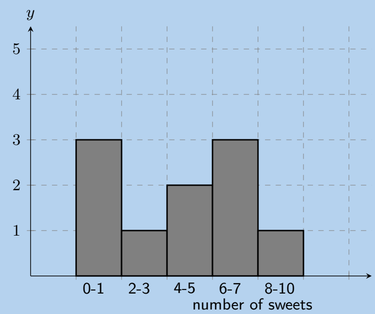
Using the ordered data set we can group the data and draw the correct histogram:



The column with the error in it was: E.

The learners used the incorrect value of **9**, when the correct value is 1.

16. A group of learners count the number of sweets they each have. This is a histogram describing the data they collected:



A cleaner knocks over their table, and all their notes land on the floor, mixed up, by accident!

Help them find which of the following data sets match the above histogram:

Data set A

1	8	4	8	8
6	1	5	7	5

Data set B

5	6	9	2	1
6	6	4	4	6

Data set C

7	2	4	1	5
1	1	7	8	6

Solution:

In order to determine which data set is correct we need to order each data set:

Data set A

1	1	4	5	5
6	7	8	8	8

Data set B

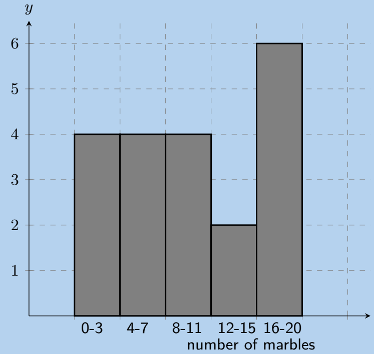
1	2	4	4	5
6	6	6	6	9

Data set C

1	1	1	2	4
5	6	7	7	8

Now we can group the data for each data set and compare the grouped data to the histogram. Doing so we find that data set C is the correct data set.

17. A group of learners count the number of marbles they each have. This is a histogram describing the data they collected:



A cat jumps onto the table, and all their notes land on the floor, mixed up, by accident!

Help them find which of the following data sets match the above histogram:

Data set A

7	13	15	13	12
13	8	14	3	15
1	7	4	11	1

Data set B

17	1	5	4	11
13	6	19	6	20
19	1	14	9	17
3	16	3	10	10

Data set C

10	3	5	5	6
5	2	1	4	3

Solution:

In order to determine which data set is correct we need to order each data set:

Data set A

1	1	3	4	7
7	8	11	12	13
13	13	14	15	15

Data set B

1	1	3	3	4
5	6	6	9	10
10	11	13	14	16
17	17	19	19	20

Data set C

1	2	3	3	4
5	5	5	6	10

Now we can group the data for each data set and compare the grouped data to the histogram. Doing so we find that data set B is the correct data set.

18. A group of **20** learners count the number of marbles they each have. This is the data they collect:

11	8	17	13	9
12	2	6	15	7
14	15	1	6	6
13	19	9	6	19

Calculate the **range** of values in the data set.

Solution:

We need to order the data set:

1	2	6	6	6
6	7	8	9	9
11	12	13	13	14
15	15	17	19	19

Now we find the maximum value in the data set:

$$\text{maximum value} = 19$$

Next we find the minimum value in the data set:

$$\text{minimum value} = 1$$

Finally, we calculate the range of the data set.

$$\begin{aligned} \text{range} &= (\text{maximum value}) - (\text{minimum value}) \\ &= 19 - 1 \\ &= 18 \end{aligned}$$

19. A group of **15** learners count the number of sweets they each have. This is the data they collect:

5	13	4	15	5
6	1	3	13	13
15	14	7	2	4

Calculate the **range** of values in the data set.

Solution:

We first need to order the data set:

1	2	3	4	4
5	5	6	7	13
13	13	14	15	15

Next we find the maximum value in the data set.

$$\text{maximum value} = 15$$

Then we find the minimum value in the data set.

$$\text{minimum value} = 1$$

Finally, we calculate the range of the data set.

$$\begin{aligned} \text{range} &= (\text{maximum value}) - (\text{minimum value}) \\ &= 15 - 1 \\ &= 14 \end{aligned}$$

20. An engineering company has designed two different types of engines for motorbikes. The two different motorbikes are tested for the time (in seconds) it takes for them to accelerate from 0 km·h^{-1} to 60 km·h^{-1}.

Test	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Bike 1	1,55	1,00	0,92	0,80	1,49	0,71	1,06	0,68	0,87	1,09
Bike 2	0,9	1,0	1,1	1,0	1,0	0,9	0,9	1,0	0,9	1,1

a) Which measure of central tendency should be used for this information?

Solution:

Mean and mode. The mean will give us the average acceleration time, while the mode will give us the time that is most often obtained.

If we used the median we would not get any useful information as all the median tells us is what the central value is. The mean and mode provide more information about the data set as a whole.

b) Calculate the measure of central tendency that you chose in the previous question, for each motorbike.

Solution:

We first sort the data.

Bike 1: $\{0{,}68; 0{,}71; 0{,}80; 0{,}87; 0{,}92; 1{,}00; 1{,}06; 1{,}09; 1{,}49; 1{,}55\}$.

Bike 2: $\{0{,}9; 0{,}9; 0{,}9; 0{,}9; 1{,}0; 1{,}0; 1{,}0; 1{,}0; 1{,}1; 1{,}1\}$.

Next we can calculate the mean for each bike:

$$\begin{aligned}\text{mean bike 1} &= \frac{0{,}68 + 0{,}71 + 0{,}80 + 0{,}87 + 0{,}92 + 1{,}00 + 1{,}06 + 1{,}09 + 1{,}49 + 1{,}55}{10} \\ &= 1{,}02\end{aligned}$$

$$\begin{aligned}\text{mean bike 2} &= \frac{0{,}9 + 0{,}9 + 0{,}9 + 0{,}9 + 1{,}0 + 1{,}0 + 1{,}0 + 1{,}0 + 1{,}1 + 1{,}1}{10} \\ &= 1{,}0\end{aligned}$$

For bike 1 the mean is 1,02 s and there is no mode, because there is no value that occurs more than once.

For bike 2 the mean is 1,0 s and there are two modes, 1,0 and 0,9.

c) Which motorbike would you choose based on this information? Take note of the accuracy of the numbers from each set of tests.

Solution:

It would be difficult to choose. Although bike 1 appears to do better than bike 2 from the mean, the data for bike 2 is less accurate than that for bike 1 (it only has 1 decimal place). If we were to calculate the mean for bike 1 using only 1 decimal place we would get 0,9 s. This would make bike 2 better. Also bike 2 produces more consistent numbers. So bike 2 would likely be a good choice, but more information or more accurate information should be obtained.

21. In a traffic survey, a random sample of 50 motorists were asked the distance they drove to work daily. This information is shown in the table below.

Distance (km)	Count
$0 < d \leq 5$	4
$5 < d \leq 10$	5
$10 < d \leq 15$	9
$15 < d \leq 20$	10
$20 < d \leq 25$	7
$25 < d \leq 30$	8
$30 < d \leq 35$	3
$35 < d \leq 40$	2
$40 < d \leq 45$	2

a) Find the approximate mean of the data.

Solution:

To find the approximate mean we need to use the central value from each group. We are told that 50 motorists were surveyed and so the total number of data values is 50.

$$\begin{aligned}\text{mean} &= \frac{4(3) + 5(8) + 9(13) + 10(18) + 7(23) + 8(28) + 3(33) + 2(38) + 2(43)}{50} \\ &= 19{,}9\end{aligned}$$

b) What percentage of drivers had a distance of

i. less than or equal to 15 km?

ii. more than 30 km?

iii. between 16 km and 30 km?

Solution:

i. The first three groups all drive less than or equal to 15 km. We can add up the number of drivers in these three groups and then divide this by the total number of drivers to find the percentage of drivers: $\frac{18}{50} \times 100 = 38\%$.

ii. The last three groups all drive more than 30 km. We can add up the number of drivers in these three groups and then divide this by the total number of drivers to find the percentage of drivers: $\frac{6}{50} \times 100 = 12\%$.

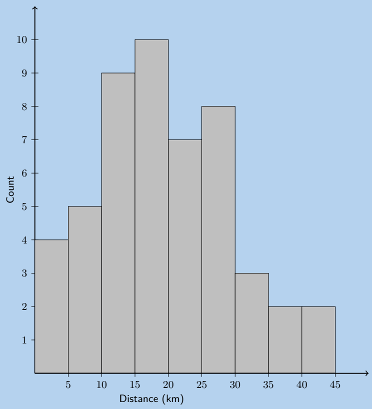
iii. The middle three groups fall into this range. We can add up the number of drivers in these three groups and then divide this by the total number of drivers to find the percentage of drivers: $\frac{25}{50} \times 100 = 50\%$.

Note that the three percentages we have just calculated all add up to 100%.

c) Draw a histogram to represent the data.

Solution:

We are given the groupings and the counts for each group. So we can draw the following histogram to represent the data:



22. A company wanted to evaluate the training programme in its factory. They gave the same task to trained and untrained employees and timed each one in seconds.

Trained	121	137	131	135	130
	128	130	126	132	127
	129	120	118	125	134
Untrained	135	142	126	148	145
	156	152	153	149	145
	144	134	139	140	142

a) Find the medians and quartiles for both sets of data.

Solution:

First order the data sets for both trained and untrained employees.

Trained: $118, 120, 121, 125, 126, 127, 128, 129, 130, 130, 131, 132, 134, 135, 137$.

Untrained: $126, 134, 135, 139, 140, 142, 142, 144, 145, 145, 148, 149, 152, 153, 156$.

There are 15 values in each data set.

Using the percentile formula with $n = 15$, we can find the rank of the 25^{th}, 50^{th} and 75^{th} percentiles:

$$\begin{aligned} r_{25} &= \frac{25}{100}(15-1)+1 \\ &= 4{,}5 \\ r_{50} &= \frac{50}{100}(15-1)+1 \\ &= 8 \\ r_{75} &= \frac{75}{100}(15-1)+1 \\ &= 11{,}5 \end{aligned}$$

For the 25^{th} percentile the rank is 4,5, which is between the fourth and fifth values. For the 50^{th} percentile (the median) the rank is 8. Therefore the median lies at the eighth value. For the 75^{th} percentile the rank is 11,5, meaning between the eleventh and 12$^{\text{th}}$ values.

For the trained employees we get:

25$^{\text{th}}$ percentile: 125,5; median: 129; 75$^{\text{th}}$ percentile: 131,5.

For the untrained employees we get:

25$^{\text{th}}$ percentile: 139,5; median: 144; 75$^{\text{th}}$ percentile: 148,5.

b) Find the interquartile range for both sets of data.

Solution:

Interquartile range for the trained employees: $Q_3 - Q_1 = 6$.

Interquartile range for the untrained employees: $Q_3 - Q_1 = 9$.

c) Comment on the results.

Solution:

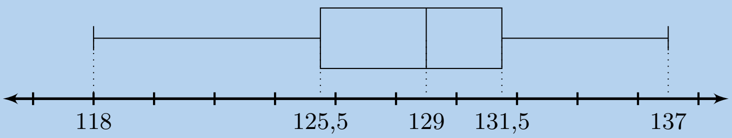
The median of the untrained employees is higher than that of the trained employees. Also the untrained employees have a larger interquartile range than the trained employees. There is some evidence to suggest that the training programme may be working.

d) Draw a box-and-whisker diagram for each data set to illustrate the five number summary.

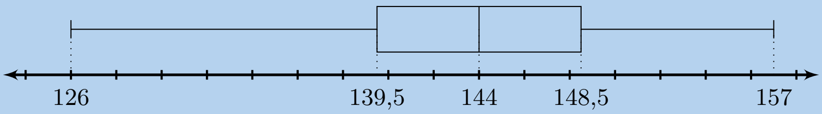
Solution:

A box-and-whisker plot shows the five number summary. The box shows the interquartile range (the distance between $Q1$ and $Q3$). A line inside the box shows the median. The lines extending outside the box (the whiskers) show where the minimum and maximum values lie.

Trained employees:



Untrained employees:



23. A small firm employs 9 people. The annual salaries of the employees are:

R 600 000	R 250 000	R 200 000
R 120 000	R 100 000	R 100 000
R 100 000	R 90 000	R 80 000

a) Find the mean of these salaries.

Solution:

$$\begin{aligned} \text{mean} &= \frac{600\ 000 + 250\ 000 + 200\ 000 + 120\ 000 + 3(100\ 000) + 90\ 000 + 80\ 000}{9} \\ &= \frac{1\ 640\ 000}{9} \\ &= \text{R } 182\ 222{,}22 \end{aligned}$$

b) Find the mode.

Solution:

The mode is R 100 000 (this value occurs 3 times in the data set).

c) Find the median.

Solution:

First order the data. To make the numbers easier to work with we will divide each one by 100 000.

The ordered set is $\{80; 90; 100; 100; 100; 120; 200; 250; 600\}$.

The median is at position 5 and is R 100 000.

d) Of these three figures, which would you use for negotiating salary increases if you were a trade union official? Why?

Solution:

Either the mode or the median.The mean is skewed (shifted) by the one salary of R 600 000. The mode gives us a better estimate of what the employees are actually earning. The median also gives us a fairly accurate representation of what the employees are earning.

24. The stem-and-leaf diagram below indicates the pulse rate per minute of ten Grade 10 learners.

$$\begin{array}{r|llll} 7 & 8 & & & \\ 8 & 1 & 3 & 5 & 5 \\ 9 & 0 & 1 & 1 & \\ 10 & 3 & 5 & & \end{array}$$

Key: $7|8 = 78$

a) Determine the mean and the range of the data.
Solution:
The data set is $\{78; 81; 83; 85; 85; 90; 91; 91; 103; 105\}$.

$$\begin{aligned} \bar{x} &= \frac{x_1 + x_2 + ...x_n}{N} \\ &= \frac{892}{10} \\ \bar{x} &= 89{,}2 \end{aligned}$$

$$\begin{aligned} \text{range} &= \text{maximum} - \text{minimum} \\ &= 105 - 78 \\ \text{range} &= 27 \end{aligned}$$

The mean and range are 89,2 and 27 respectively.

b) Give the five-number summary and create a box-and-whisker plot for the data.
Solution:

$$\begin{aligned} r &= \frac{p}{100}(n-1) + 1 \\ r_{25} &= \frac{25}{100}(10-1) + 1 \\ &= 3{,}25 \end{aligned}$$

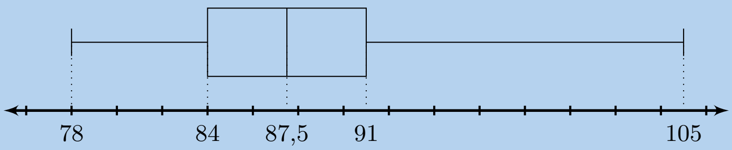
$$\begin{aligned} \therefore Q_1 &= \frac{83 + 85}{2} \\ &= 84 \end{aligned}$$

$$\begin{aligned} r_{50} &= \frac{50}{100}(10-1) + 1 \\ &= 5{,}5 \end{aligned}$$

$$\begin{aligned} \therefore Q_2 &= \frac{85 + 90}{2} \\ &= 87{,}5 \end{aligned}$$

$$\begin{aligned} r_{75} &= \frac{75}{100}(10-1) + 1 \\ &= 7{,}75 \\ \therefore Q_3 &= \frac{91 + 91}{2} \\ &= 91 \end{aligned}$$

The five-number summary is: 78; 84; 87,5; 91; 105.
Using this we can draw the box-and-whisker plot.



25. The following is a list of data: $3; 8; 8; 5; 9; 1; 4; x$
In each separate case, determine the value of x if the:

a) range = 16

Solution:

The set is: $\{1; 3; 4; 5; 8; 8; 9; x\}$. We have ordered all the numbers and then added x on the end as we do not know the value of x.

We are told that the range must be equal to 16. If $x < 9$ the range would be $9 - 1 = 8$. Therefore $x > 9$ and so x must be the maximum value.

$$\begin{aligned} \text{range} &= \text{maximum} - \text{minimum} \\ 16 &= x - 1 \\ \therefore x &= 17 \end{aligned}$$

b) mode = 8

Solution:

8 is already the mode if we exclude x, to maintain this x is any integer with $x \neq \{1; 3; 4; 5; 9\}$.

c) median = 6

Solution:

First consider the set without x: $\{1; 3; 4; 5; 8; 8; 9\}$. The median in this set is 5 (there is an odd number of values in the set and the median lies at position 4).

Next we consider the full set: $\{1; 3; 4; 5; 8; 8; 9; x\}$ There is an even number of values (8) in the full set. Therefore the median must lie between the fourth and fifth values.

Now we need to think about where x could fit into the set. x could be the fourth value, the fifth value or somewhere else in the set. If x is either the fourth or fifth value we will get the same median.

First try the case where x is the fourth or fifth value:

$$\begin{aligned} 6 &= \frac{5 + x}{2} \\ x + 5 &= 12 \\ \therefore x &= 7 \end{aligned}$$

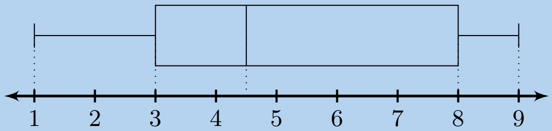
Next we check the case where x is not the fourth or fifth value. In this case the median is $\frac{5+8}{2}$ = 6,5. Therefore we can say that $x = 7$.

d) mean = 6

Solution:

$$\begin{aligned} \bar{x} &= \frac{x_1 + x_2 + ... x_n}{N} \\ 6 &= \frac{x + 38}{8} \\ x + 38 &= 48 \\ \therefore x &= 10 \end{aligned}$$

e) box-and whisker plot



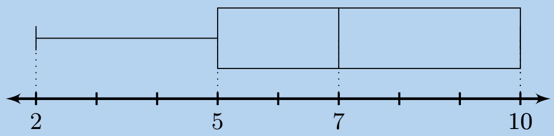
Solution:

In part a we found that if the range is 8, then $x < 9$. The range here is 8, therefore $x < 9$.

Therefore we will use the median to help us find x. The median on the box-and-whiskers plot is 4,5. From our reasoning in part c we know that this means x is either the fourth or fifth value. So we can calculate x as follows:

$$\begin{aligned} 4{,}5 &= \frac{5 + x}{2} \\ x + 5 &= 9 \\ \therefore x &= 4 \end{aligned}$$

26. Write down one list of numbers that satisfies the box-and-whisker plot below:



Solution:

From the box-and-whisker plot we get the five number summary: $2; 5; 7; 10; 10$. Note that the third quartile is also the maximum value in this case.

From this we can state that the data set must have a minimum value of 2 and a maximum value of 10.

The data set can contain any numbers in the range $2 \leq x \leq 10$ such that the first quartile is 5, the median is 7 and the third quartile is 10. There is also no restriction on how many values are in the data set.

One possible set that satisfies this set of numbers is $\{2; 5; 7; 10; 10\}$. You can check that this set works by calculating the quartiles.

27. Given ϕ (which represents the golden ratio) to 20 decimal places: 1,61803398874989484820

a) For the first 20 decimal digits of (Φ), determine the:

i. median

ii. mode

iii. mean

Solution:

i. The ordered set is: $\{0; 0; 1; 2; 3; 3; 4; 4; 4; 6; 7; 8; 8; 8; 8; 8; 8; 9; 9; 9\}$

$$\begin{aligned} r_{50} &= \frac{50}{100}(20-1)+1 \\ &= 10{,}5 \end{aligned}$$

$$\begin{aligned} \text{median} &= \frac{6+7}{2} \\ &= 6{,}5 \end{aligned}$$

ii. The mode is 8.

iii.

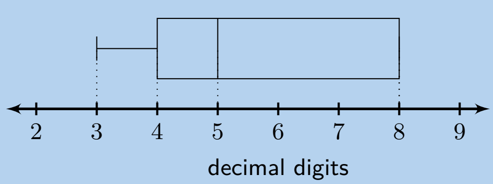
$$\begin{aligned} \bar{x} &= \frac{x_1+x_2+\ldots x_n}{N} \\ &= \frac{109}{20} \\ \bar{x} &= 5{,}45 \end{aligned}$$

b) If the mean of the first 21 decimal digits of (Φ) is 5,38095 determine the 21$^{\text{st}}$ decimal digit.

Solution:

$$\begin{aligned} \bar{x} &= \frac{x_1+x_2+\ldots x_n}{N} \\ 5{,}38095 &= \frac{109+\phi_{21}}{21} \\ \phi_{21} &= 21(5{,}38095)-109 \\ \phi_{21} &= 4 \end{aligned}$$

c) Below is a box-and-whisker plot of the 21$^{\text{st}}$ - 27$^{\text{th}}$ decimal digits. Write down one list of numbers that satisfies this box-and-whisker plot.



Solution:

From the box-and-whisker plot we get the five number summary: $3; 4; 5; 8; 8$. Note that the third quartile is also the maximum value in this case.

From this we can state that the data set must have a minimum value of 3 and a maximum value of 8.

The data set can contain any numbers in the range $3 \leq x \leq 8$ such that the first quartile is 4, the median is 5 and the third quartile is 8. However, we know that the data set consists of the 21$^{\text{st}}$ - 27$^{\text{th}}$ decimal digits of Φ and so the data set must contain 7 values.

The median will be at the fourth position and so the fourth number in the set is 5. The first quartile will lie between the second and third values while the third quartile will lie between the fifth and sixth values.

Let the data set be: $\{3; x; y; 5; a; b; 8\}$

The first quartile is:

$$\begin{aligned} 4 &= \frac{x+y}{2} \\ 8 &= x+y \end{aligned}$$

x and y can be any integers that add up to 8. However x and y must be greater than or equal to 3 and less than or equal to 5. Therefore the possible values are: 3 and 5 or 4 and 4.

The third quartile is:

$$\begin{aligned} 8 &= \frac{a+b}{2} \\ 16 &= a+b \end{aligned}$$

a and b can be any integers that add up to 16. However a and b must be greater than or equal to 5 and less than or equal to 8. Therefore the only possible values are: 8 and 8.

There are two possible sets: $\{3; 3; 5; 5; 8; 8; 8\}$ or $\{3; 4; 4; 5; 8; 8; 8\}$.

28. There are 14 men working in a factory. Their ages are : $22; 25; 33; 35; 38; 48; 53; 55; 55; 55; 55; 56; 59; 64$

a) Write down the five number summary.

Solution:

$$\begin{aligned} r &= \frac{p}{100}(n-1)+1 \\ r_{25} &= \frac{25}{100}(14-1)+1 \\ &= 4{,}25 \end{aligned}$$

$$\begin{aligned} \therefore Q_1 &= \frac{35+38}{2} \\ &= 36{,}5 \end{aligned}$$

$$\begin{aligned} r_{50} &= \frac{50}{100}(14-1)+1 \\ &= 7{,}5 \end{aligned}$$

$$\begin{aligned} \therefore Q_2 &= \frac{53+55}{2} \\ &= 54 \end{aligned}$$

$$\begin{aligned} r_{75} &= \frac{75}{100}(14-1)+1 \\ &= 10{,}75 \end{aligned}$$

$$\begin{aligned} \therefore Q_3 &= \frac{55+55}{2} \\ &= 55 \end{aligned}$$

The five number summary is: 22; 36,5; 50; 55; 64

b) If 3 men had to be retrenched, but the median had to stay the same, show the ages of the 3 men you would retrench.

Solution:

Retrenching 3 men will leave 11 men. For an odd numbered set the median must be the same as one of the ages. None of the men are 54 years.

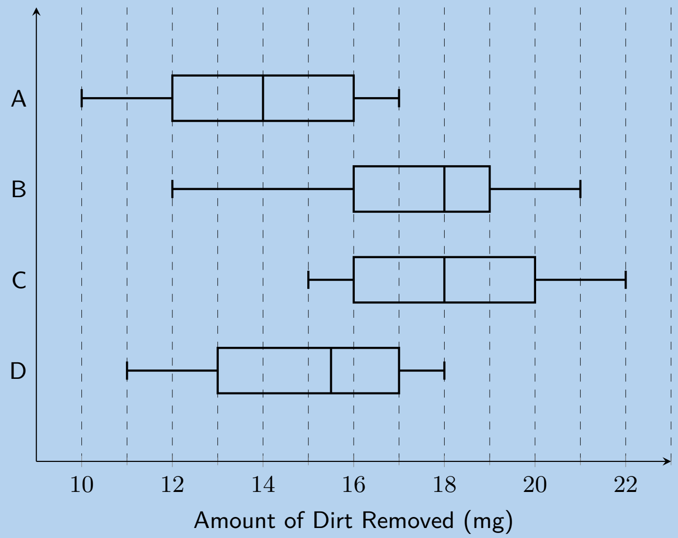
Therefore no men can be retrenched to keep the median the same.

c) Find the mean age of the men in the factory using the original data.

Solution:

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 + ...x_n}{N}\\ &= \frac{597}{14}\\ \therefore \bar{x} &= 42{,}643\end{aligned}$$

29. The example below shows a comparison of the amount of dirt removed by four brands of detergents (brands A to D).



a) Which brand has the biggest range, and what is this range?

Solution:

A: range $= 17 - 10 = 7$

B: range $= 21 - 12 = 9$

C: range $= 22 - 15 = 7$

D: range $= 18 - 11 = 7$

B has the biggest range. The range is 9.

b) For brand C, what does the number 18 mg represent?

Solution:

18 mg represents the median.

c) Give the interquartile range for brand B.

Solution:

$$\begin{aligned}\text{interquartile range} &= Q_3 - Q_1\\ &= 19 - 16\\ &= 3\end{aligned}$$

d) Which brand of detergent would you buy? Explain your answer.

Solution:

We need to compare several values to help us decide. These values are shown in the table below.

Brand	Minimum value	Maximum value	Range	Interquartile range
A	10	17	7	4
B	12	21	9	3
C	15	22	7	4
D	11	18	7	4

From this we see that brand C has the highest minimum value. Brand B has the smallest interquartile range but the largest range. It is possible that the minimum value for brand B is an outlier which would make brand B a better choice than brand C.

Considering all the data available it would be hard to choose between brand C and brand B. We can however say that brands A and D are not very good choices as they both have low minimum and maximum values.

Since brand C does not have a potential outlier this might be the best brand to choose.

For more exercises, visit www.everythingmaths.co.za and click on 'Practise Maths'.

1. 2GNM 2. 2GNN 3. 2GNP 4. 2GNQ 5. 2GNR 6. 2GNS
7. 2GNT 8. 2GNV 9. 2GNW 10. 2GNX 11. 2GNY 12. 2GNZ
13. 2GP2 14. 2GP3 15. 2GP4 16. 2GP5 17. 2GP6 18. 2GP7
19. 2GP8 20. 2GP9 21. 2GPB 22. 2GPC 23. 2GPD 24. 2GPF
25. 2GPG 26. 2GPH 27. 2GPJ 28. 2GPK 29. 2GPM

www.everythingmaths.co.za **m.everythingmaths.co.za**