

SESSION 10: DATA HANDLING

Key Concepts

In this session we will focus on summarising what you need to know about:

- Mean
- Mode
- Quartiles
- Range
- The Inter-quartile Range (IQR)
- Five Number Summaries
- Box and Whisker Plots
- Standard deviation using a table
- Calculator programmes to calculate standard deviation

X-planation

Mean

The mean of a set of data is the average. To get the mean, you add the scores and divide by the number of scores.

Mode

This is the most frequently occurring score.

Quartiles

Quartiles are measures of dispersion around the median which is a good measure of central tendency. The median divides the data into two halves. The lower and upper quartiles further subdivide the data into quarters.

There are three quartiles:

The Lower Quartile (Q_1): This is the median of the lower half of the values.

The Median (M or Q_2): This is the value that divides the data into halves.

The Upper Quartile (Q_3): This is the median of the upper half of the values.

If there is an **odd** number of data values in the data set, then the specific quartile will be a value in the data set. If there is an **even** number of data values in the data set, then the specific quartile will not be a value in the data set. A number which will serve as a quartile will need to be inserted into the data set (the average of the two middle numbers).

Range

The range is the difference between the largest and the smallest value in the data set. The bigger the range, the more spread out the data is.

The Inter-quartile Range (IQR)

The difference between the lower and upper quartile is called the inter-quartile range.

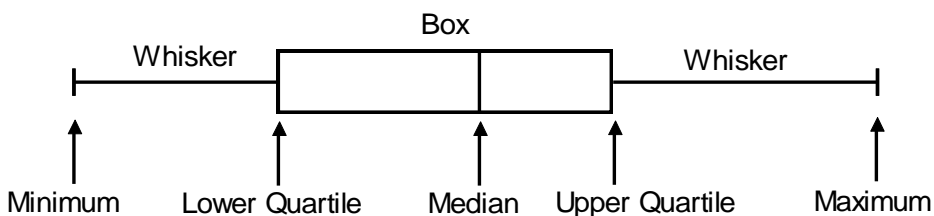
Five Number Summaries

The Five Number Summary uses the following measures of dispersion:

- Minimum: The smallest value in the data
- Lower Quartile: The median of the lower half of the values
- Median: The value that divided the data into halves
- Upper Quartile: The median of the upper half of the values
- Maximum: The largest value in the data

Box and Whisker Plots

A Box and Whisker Plot is a graphical representation of the Five Number Summary.



Standard deviation using a table

The standard deviation (SD) can be determined by using the following formula:

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Calculator programmes to calculate standard deviation

CASIO fx-82ES PLUS:

MODE

2 : STAT

1 : 1 – VAR

Enter the data points: push = after each data point

AC

SHIFT

www.learnxtra.co.za

STAT

5: VAR

3 : $x\sigma n$

push = to get standard deviation

SHARP DAL

MODE 1 =

Enter data points: push STO 2 M+ after each data point

RCL 6 to get standard deviation

X-ample Questions

Question 1

The results for Argentina for the past 14 World Cup tournaments are recorded in the table below.

World Cup tournament	Matches played	Wins	Draws	Losses	Goals for	Goals against
2006 Germany	5	3	2	0	11	3
2002 Korea/Japan	3	1	1	1	2	2
1998 France	5	3	1	1	10	4
1994 USA	4	2	0	2	8	6
1990 Italy	7	2	3	2	5	4
1986 Mexico	7	6	1	0	14	5
1982 Spain	5	2	0	3	8	7
1978 Argentina	7	5	1	1	15	4
1974 Germany	6	1	2	3	9	11
1966 England	4	2	1	1	4	2
1962 Chile	3	1	1	1	2	3
1958 Sweden	3	1	0	2	5	10
1934 Italy	1	0	0	1	2	3
1930 Uruguay	5	4	0	1	18	9

Source: [www.2010](http://www.2010FifaWorldCup.com) Fifa World Cup:Statistics – MediaClubSouthAfrica.com

- Determine the quartiles for the matches played. (3)
- Determine the quartiles for the wins. (3)
- Determine the quartiles for the goals scored against Argentina. (3)
- Draw box and whisker plots for the matches played by Argentina, the wins and the goals scored against Argentina. (15)
- By referring to the box and whisker plots, comment on the distribution of the data. (6)

(f) Calculate the mean for the number of matches played over the 14 tournaments. (2)

(g) Complete the following table:

Matches played (x)	$x - \bar{x}$	$(x - \bar{x})^2$
5		
3		
5		
4		
7		
7		
5		
7		
6		
4		
3		
3		
1		
5		
		$\sum (x - \bar{x})^2 =$

(h) Use the table to calculate the standard deviation for the matches played. (3)

(i) Using your calculator, calculate the standard deviation for the goals for and the goals against Argentina. Comment on your findings. (4)

Question 2

The ages of the final 23 players selected by coach Carlos Perreira to play for Bafana Bafana in the 2010 FIFA World Cup are provided below.

Position	Player	Age
1	Shu-Aib Walters	28
2	Siboniso Gaxa	26
3	Tshepo Masilela	25
4	Aaron Moekoena (captain)	29
5	Lucas Thwala	28
6	Macbeth Sibaya	32
7	Lance Davids	25
8	Siphiwe Tshabalala	25
9	Katlego Mphela	25
10	Steven Pienaar	28



11	Teko Modise	27
12	Reneilwe Letsholonyane	28
13	Kagisho Dikgacoi	25
14	Matthew Booth	33
15	Bernard Parker	24
16	Itumeleng Khune	22
17	Surprise Moriri	30
18	Siyabonga Nomvetho	32
19	Anele Ngcongca	22
20	Bongani Khumalo	23
21	Siyabonga Sangweni	28
22	Moeneeb Josephs	30
23	Thanduyise Khuboni	24



Source: www.2010 Fifa World Cup:final squads – MediaClubSouthAfrica.com

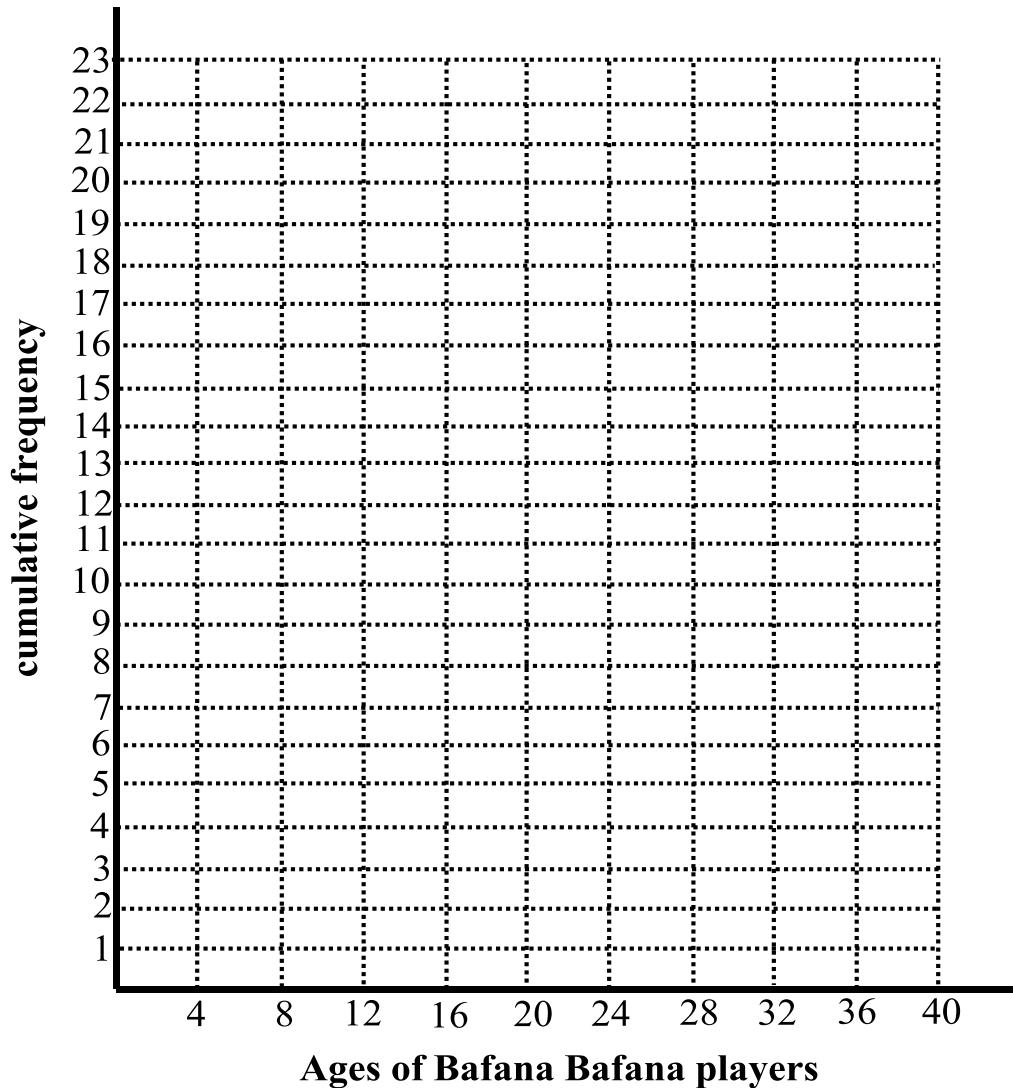
The ages of the players are to be grouped into class intervals.

(a) Complete the following table:

(2)

Class intervals (ages)	Frequency	Cumulative frequency
$16 \leq x < 20$		
$20 \leq x < 24$		
$24 \leq x < 28$		
$28 \leq x < 32$		
$32 \leq x < 36$		

- (b) On the diagram provided below, draw a cumulative frequency curve for this data. (6)



- (c) Use your graph to read off approximate values for the quartiles. (3)

Question 3

- (a) Complete the table and then use the table to calculate the standard deviation. (5)

Class intervals	Frequency (f)	Midpoint (m)	$f \times m$	$m - \bar{x}$	$(m - \bar{x})^2$	$f \times (m - \bar{x})^2$
$20 \leq x < 24$	3	22				
$24 \leq x < 28$	9	26				
$28 \leq x < 32$	8	30				
$32 \leq x < 36$	3	34				
			$\bar{x} =$			

www.learnxtra.co.za

- (b) Hence calculate the standard deviation using the table. (2)
- (c) Now use your calculator to verify your answer. (2)