SESSION 14: DATA REPRESENTATION 2

Key Concepts

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

- Analysing data.
- Measures of central tendency
- Measures of spread

X-planation

ANALYSING THE DATA

1. INTRODUCTION

Once we have our data sorted and displayed it is important to analyse it. In other words, use figures to determine exactly what is happening and draw accurate conclusions from them.

2. MEASURES OF CENTRAL TENDENCY

"Measures of central tendency" are values that tell us what is happening in the middle of the data. If we know the central/middle value, then it is possible to compare the other values in the set to that middle/central one. It is also possible to use the central/middle value to compare different sets of data.

There are three measures of central tendency that we make use of:

- The Mean
  When most people talk about the average this is in fact the mean.
  The mean is calculated by adding all the values in the data set together and then dividing by the number of values in the data set.

  Example: Calculate the mean of 4, 7, 15, 8, 9, 7, 13
  4+7+15+8+9+7+13 = 63 ÷ 7 = 9

- The Median
  The median is the value that literally lies in the middle of all the values. You must first order the data before being able to determine the median accurately.

  Example: What is the median of this set of data 4, 7, 15, 8, 9, 7, 13?
  4, 7, 7, 8, 9, 13, 15
  Median is 8

  What is the median of this set of data 5, 7, 8, 8, 10, 11, 14, 14
  The median is between 8 and 10

$$\frac{8+10}{2} = 9$$

- The Mode
  The mode is the average that indicates the value that has occurred the most in the given set. Note there are times when there is no mode and others where there are more than one mode.

  Example: Calculate the mode of 4, 7, 15, 8, 9, 7, 13
  Mode is 7

3. MEASURES OF SPREAD

In some cases it is not enough to know the average of the data; for this reason we make use of the "Measures of Spread". These values reflect the spread of the data, i.e. how far apart do all the values lie from each other.

The measures of spread that we use are:

- Range
  The range takes into account the highest value and the lowest value in the data set. In order to calculate the range you need to subtract these two values.

  Example: Calculate the range of 4, 7, 15, 8, 9, 7, 13
  Range = 15 – 4 = 11
  The range is not a good measure of spread if there are outliers in the data (outliers are values that are extremely different to all other values given).

- Quartiles
  The quartiles divide the set of data into quarters or 25% of the data. You must first order the data before being able to determine the quartiles accurately. The median is the second quartile Q2. The first quartile Q1 is the median of the lower half of the set. The third quartile Q3 is the median of the upper half of the set.

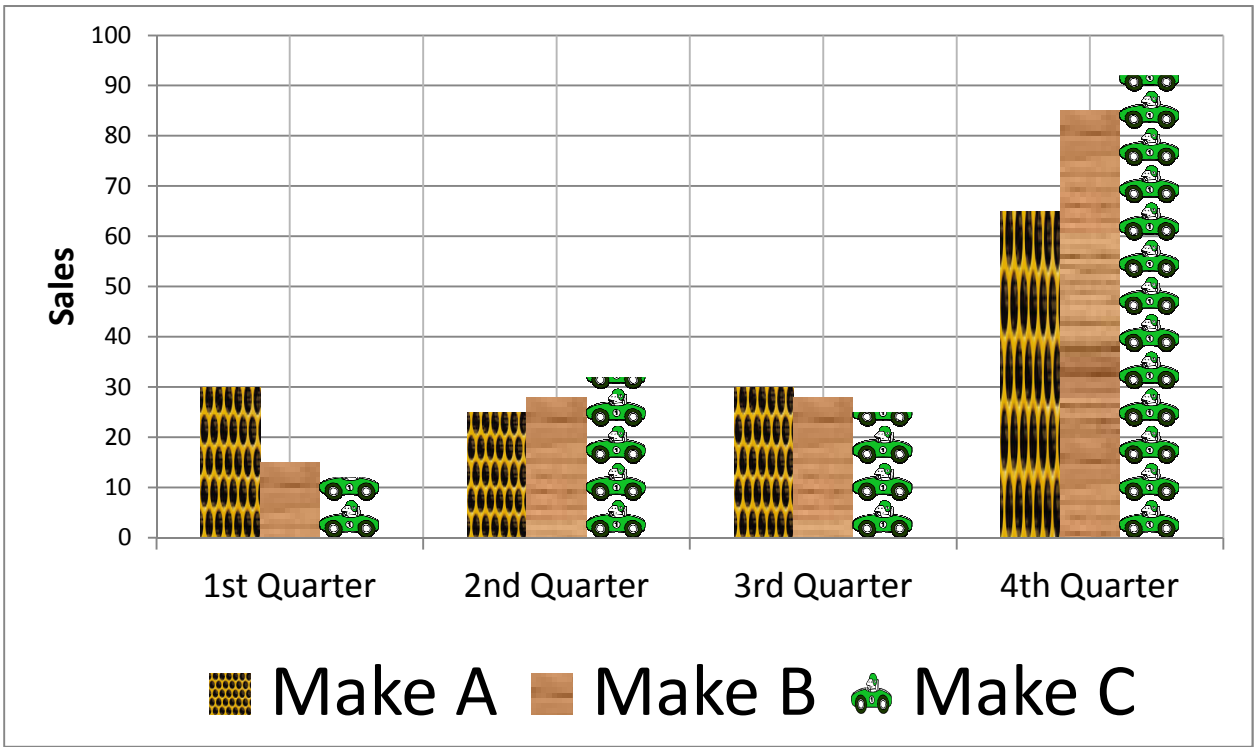
  Example: What are the quartiles of this set of data 4, 7, 15, 8, 9, 7, 13?
  4, 7, 7, 8, 9, 13, 15
  Q1 = 7, Q2 = 8 and Q3 = 13

- The Inter-quartile range (IQR) = Q3 – Q1 This is a very useful spread because it eliminates the outliers and only focuses on 50% of the data around the median.

X-ample Questions

Question 1

The graph below shows the car sales of three makes of cars for the quarters over the last year.



a) How many units of Make A were sold in the second quarter? (2)

b) Which make of car had the lowest number of sales in the third quarter? (2)

c) Compare the sales in the first and last quarters. Suggest a reason for this trend. (2)

Question 2

The following tables show the number of vehicles passing 2 intersections over a ten hour period; from the morning till afternoon.

Intersection A

Time (H)	Number of Cars
7-8	351
8-9	245
9-10	102
10-11	100
11-12	235
12-13	245
13-14	96
14-15	155
15-16	211
16-17	332

Intersection B

Time (H)	Number of Cars
7-8	455
8-9	325
9-10	92
10-11	107
11-12	344
12-13	378
13-14	89
14-15	289
15-16	325
16-17	468

d) Calculate the mean number of cars passing both intersections per hour. (6)
e) Calculate the median number of cars for intersection A. (4)
f) Which intersection is busier between and during the first hour of the day? (1)
g) State the mode of intersection B. (1)
h) If the city council only has a budget to install traffic lights at one of the intersections, which would you recommend? Give a reason for your answer. (2)

Question 3

52 learners wrote the Grade 11 Geography examination. The ages (in years) of a sample of 16 of these learners are as follows:

16 16 16 16 17 17 17 17 17 18 18 18 18 19 19 19

a) What age sample is the mode? (1)
b) Determine the median age of the sample of learners. (1)
c) Calculate the mean age of the sample of learners. (3)

The Geography examination marks, expressed as a percentage, of the 52 learners were recorded as follows:

54 67 83 34 49 56 78 89 90 79 20 49 50
70 89 57 27 48 56 65 70 22 98 89 29 56
47 95 49 67 89 48 46 89 63 75 45 50 58
73 67 45 76 70 38 46 37 47 36 38 99 55

d) Determine the range of marks (3)
e) The NCS (National Curriculum Statement) requires that results be expressed in terms of seven performance levels rather than percentages. (80-100, 70-79, 60-69, 50,59, 40-49, 30-39, and 0-29) As a result the geography teacher needs to work out the number of learners per performance level. Complete a frequency table to work out the number of learners per performance level. (7)

X-ercises

The table below shows the daily visitors at the Fekile Game Reserve from 17 to 31 December 2007.

December	Visitors
17	434
18	383
19	398
20	402
21	434
22	503
23	523
24	434
25	0
26	804
27	650
28	660
29	679
30	503
31	705

a) How many visited the game reserve on 20 December 2007? (1)

b) On which day was the game reserve visited by the most people? (1)

c) Why do you think there were no visitors to the game reserve on 25 December 2007? (1)

d) Calculate the total number of visitors for this period. (2)

e) Calculate the mean number of visitors for this period. (2)

f) Rearrange the number of visitors to the game reserve from the minimum to the maximum. (1)

g) Calculate the range of the above data. (2)

h) What is the mode of the data? (1)

i) Calculate the median. (1)

j) Calculate the inter-quartile range. (4)

k) Calculate the 60^{th} percentile. (4)