STATISTICS II   7 OCTOBER 2013

Lesson Description

In this lesson we:

- Examine cumulative frequency tables and curves

Key Concepts

Measures of Central Tendency

A measure of central tendency is a value that represents a typical item in a set of data.

Mean

Mean = sum of all the values ÷ total number of values
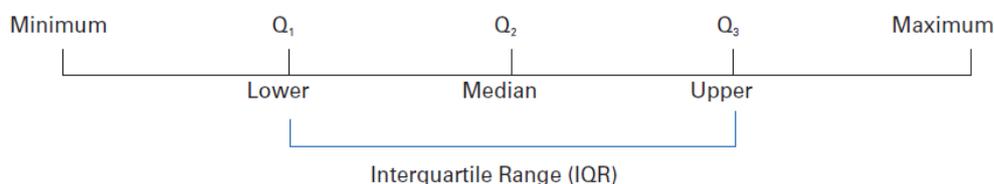
Median

Median (M) is the middle value of a data set after ordering the data. If there is an even number of values, there are two middle values so the median will be between those two values.

Mode

The mode is the value that occurs most often.

Measures of Dispersion

We can use the median to divide data into **quartiles** to show how the data is spread out around the median.

An **outlier** is an observation that lies an abnormal distance from other values in a set of data. Outliers will skew data sets. An outlier can be defined as a value that is either:
$1,5 \times IQR$ (or more) $< Q_1$
$1,5 \times IQR$ (or more) $> Q_3$ where IQR is the interquartile range

Cumulative frequency

A **cumulative frequency table** is a table in which we keep what is commonly known as a 'running total'.

Example

Mrs Mhkize records the marks out of 10 for a test that her class of 32 learners wrote. The marks obtained are given in the following frequency table.

Mark	Frequency (number of times)	Cumulative total (running total)	Cumulative frequency	
1	2	2	2	
2	3	2 + 3	5	
3	1	5 + 1	6 ←	Six learners achieved 3 marks or less
4	6	6 + 6	12	
5	8	12 + 8	20	
6	3	20 + 3	23	
7	4	23 + 4	27	
8	2	27 + 2	29 ←	29 learners achieved 8 marks or less
9	2	29 + 2	31	
10	1	31 + 1	32	
TOTAL	32 ←			

Cumulative frequency curve

The information in a cumulative frequency table can now be extended to a graphical representation, which is called a cumulative frequency curve or ogive (pronounced o-jive). The ogive is an S-shaped curve for data that has been grouped into class intervals. Note that the steepest section of the curve indicates the greatest number of values in the particular class interval. This will be the modal class interval. We can use the curve to find estimates of the median, and upper and lower quartiles of the data

Distribution of Promotion Marks for a Grade 11 Class
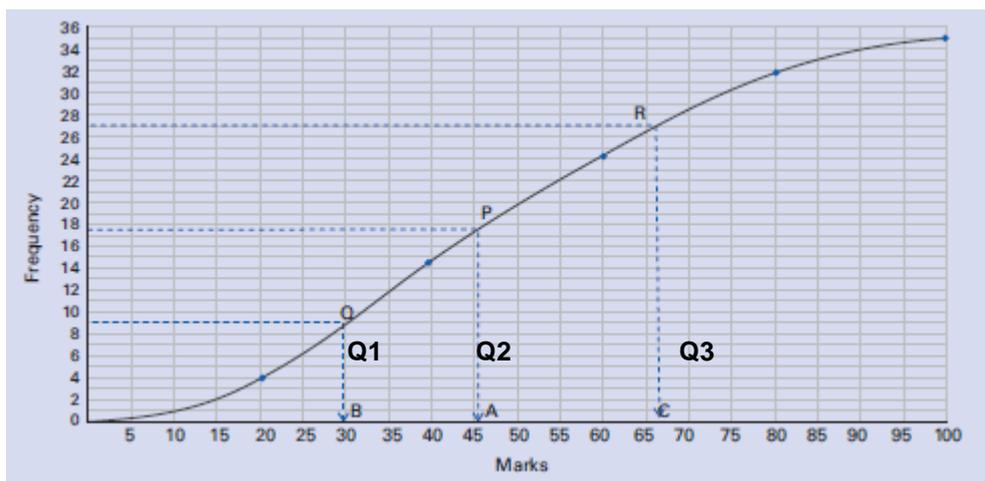
(Clever Keeping Maths Simple. Grade 11 Learner's Book. Page 451)

Variance & Standard Deviation

The median is one measure of central tendency ('middleness') and the quartiles indicate the dispersion (spread) about the median.

The mean is another measure of central tendency ('middleness') and the standard deviation gives us the dispersion (spread) about the mean. The standard deviation is denoted by the symbol σ.

The symbol for variance is $σ^2$. Thus, $σ = \sqrt{variance}$, which is our standard deviation.

The variance is given by the formula $σ^2 = \dfrac{\sum(x-\bar{x})^2}{n}$ where n is the number of data items

The standard deviation can be calculated manually using a table, or it can very easily be determined by using a calculator in STAT mode..

Questions

Question 1

(Clever Keeping Maths Simple. Grade 11 Learner's Book. Ex 13.4 Pg 453 Question 1)

Tariffs for South African toll roads increased for light motor vehicles on 1 April 2010. The minimum toll fee was R6,00 at Izotsha, while the highest was R65 at Swartruggens. The fees for the 30 toll gates in South Africa have been grouped in intervals of R10 and listed in the table that follows.

Tariffs in rands	Frequency	Cumulative frequency	Points to plot
$0 < x \le 10$	8		
$10 < x \le 20$	3		
$20 < x \le 30$	6		
$30 < x \le 40$	7		
$40 < x \le 50$	4		
$50 < x \le 60$	1		
$60 < x \le 70$	1		
TOTAL			

(Source: The Star 01-04-10)

a) Complete the given table.

b) Draw a cumulative frequency curve to represent this set of data.

c) Use your graph to determine an estimate of the:

    i) median

    ii) lower quartile

    iii) upper quartile, showing where all readings are taken and working correct to one decimal place.

d) Use your graph to determine approximately how many toll gates levy a charge of R34 or less.

e) Give the modal class of the given data set.

Question 2

(Clever Keeping Maths Simple. Grade 11 Learner's Book. Ex 13.4 Pg 453 Question 2)

The temperatures of a number of cities around the world on 26 March 2010 varied considerably. The lowest listed temperature was 2 °C at Toronto, while the highest was 34 °C in Mecca. The temperatures of 31 cities have been grouped in intervals of five degrees centigrade and listed in the following table:

Temperatures in °C	Frequency	Cumulative frequency	Points to plot
$0 < x \le 5$	1		
$5 < x \le 10$	2		
$10 < x \le 15$	7		
$15 < x \le 20$	8		
$20 < x \le 25$	5		
$25 < x \le 30$	6		
$30 < x \le 35$	2		
TOTAL			

a)   Complete the given table.

b)   Draw a cumulative frequency curve to represent this set of data.

c)   Use your graph to determine an estimate of the:

   i)   median

   ii)   lower quartile

   iii)   upper quartile, showing where all readings are taken and working correct to one decimal place.

d)   Use your graph to determine approximately how many cities experienced temperatures of 23 °C or less.

e)   Give the modal class of the given data.

f)   Draw a box-and-whisker diagram to represent the given data set. Remember that the minimum temperature was 2 °C and the maximum was 34 °C.

Question 3

(Clever Keeping Maths Simple. Grade 11 Learner's Book. Ex 13.4 Pg 453 Question 3)

After a wet summer across many parts of South Africa in 2009, many major dam levels rose considerably. The dam with the lowest level in January 2010 was the Kalkfontein Dam, which was only 16% full. Eight dams were more than 100% full, with the Grootdraai Dam being the fullest at 105%. The levels of the 26 major dams have been grouped together in intervals of 20% and listed in the given table.

Percentage full	Frequency	Cumulative frequency	Points to plot
$0 < x \leq 20$	1		
$20 < x \leq 40$	2		
$40 < x \leq 60$	1		
$60 < x \leq 80$	3		
$80 < x \leq 100$	11		
$100 < x \leq 120$	8		
TOTAL			

(Source: The Star 21-01-2010)

a)   Complete the given table.

b)   Draw a cumulative frequency curve to represent this set of data.

c)   Use your graph to determine an estimate of the:

   i)   median

   ii)   lower quartile

   iii)   upper quartile, showing where all readings are taken and working correct to one decimal place.

d)   Use your graph to determine how many dams had a level of 72% or less. Indicate where your reading is taken by using the letter D.

e)   Give the modal class of the given data set.

Question 4
(Clever Keeping Maths Simple. Grade 11 Learner's Book. Ex 13.4 Pg 453 Question 4)

One index used to compare the levels of currencies is through the McDonald's Big Mac. The Big Mac was chosen because it is a very uniform product that is available everywhere. This index is called the Big Mac Index. It provides a quick reference to what one currency would buy in another country. In April 2009 the Big Mac cost $3,15 in the USA, but would have cost $6,67 in Iceland, $4,93 in Switzerland, $2,29 in South Africa, $1,30 in China and $0,68 in Qatar. The prices of the Big Mac in 25 countries across the world in April 2009 have been grouped in intervals of $0,50 and are listed in the given table on the next page.

a)   Complete the given table.

b)   Draw a cumulative frequency curve to represent this set of data.

c)   Use your graph to determine an estimate of the:

   i)   median

   ii)   lower quartile

   iii)   upper quartile, showing where all readings are taken and working correct to two decimal places.

d) Give an approximate value of the 89th percentile, showing where your reading is taken.

Price in dollars	Frequency	Cumulative frequency	Points to plot
$2,50 < x \leq 3,00$	14		
$3,00 < x \leq 3,50$	5		
$3,50 < x \leq 4,00$	1		
$4,00 < x \leq 4,50$	2		
$4,50 < x \leq 5,00$	1		
$5,00 < x \leq 5,50$	0		
$5,50 < x \leq 6,00$	0		
$6,00 < x \leq 6,50$	1		
$6,50 < x \leq 7,00$	1		
TOTAL			