

Paper 2 - STATISTICS

The Cumulative Frequency Curve

A cumulative frequency curve, also called the Orgive curve can be used to find the mean, quartiles (lower, upper, interquartiles, semi-interquartiles) of a given distribution. To find the cumulative frequency, find the accumulated totals and plot them against the data or score values. The cumulative frequency is formed by joining the points with a smooth curve.

Question

Answer the whole of this question on a sheet of graph paper.

The waiting time for 55 passengers at the power tools bus station in Kitwe for them to board a Lusaka bound bus on a particular day were as follows:-

Waiting time (in minutes)	$1 \leq x \leq 3$	$4 \leq x \leq 6$	$7 \leq x \leq 9$	$10 \leq x \leq 12$	$13 \leq x \leq 15$
Number of Passengers	6	11	20	13	5

- a) Calculate the estimate of the mean waiting time.
- b) Copy and complete the cumulative frequency table below.

Waiting time (in minutes)	≤ 3	≤ 6	≤ 9	≤ 12	≤ 15
Number of Passengers	6	17			55

- c) Using a horizontal scale of 2cm to represent 2 minutes for times from 0 to 15 minutes and a vertical scale of 2cm to represent 10 passengers.
Draw a smooth cumulative frequency curve.
- d) Showing your method clearly, use your graph to estimate the
 - i. Median
 - ii. Lower quartile
 - iii. Upper quartile
 - iv. Interquartile range
 - v. Semi-interquartile range
 - vi. 60th percentile
- e) Find the number of passengers who waited for more than 6 minutes.
- f) (i) If a passenger was chosen at random, find the probability that he waited for less than 9 minutes.

- g) If two passengers were chosen at random. Find the probability that they both waited for more than 12 minutes.

Relative Cumulative Frequency Table

The Table below shows a frequency table of the marks obtained by 120 pupils in a Mathematics Test.

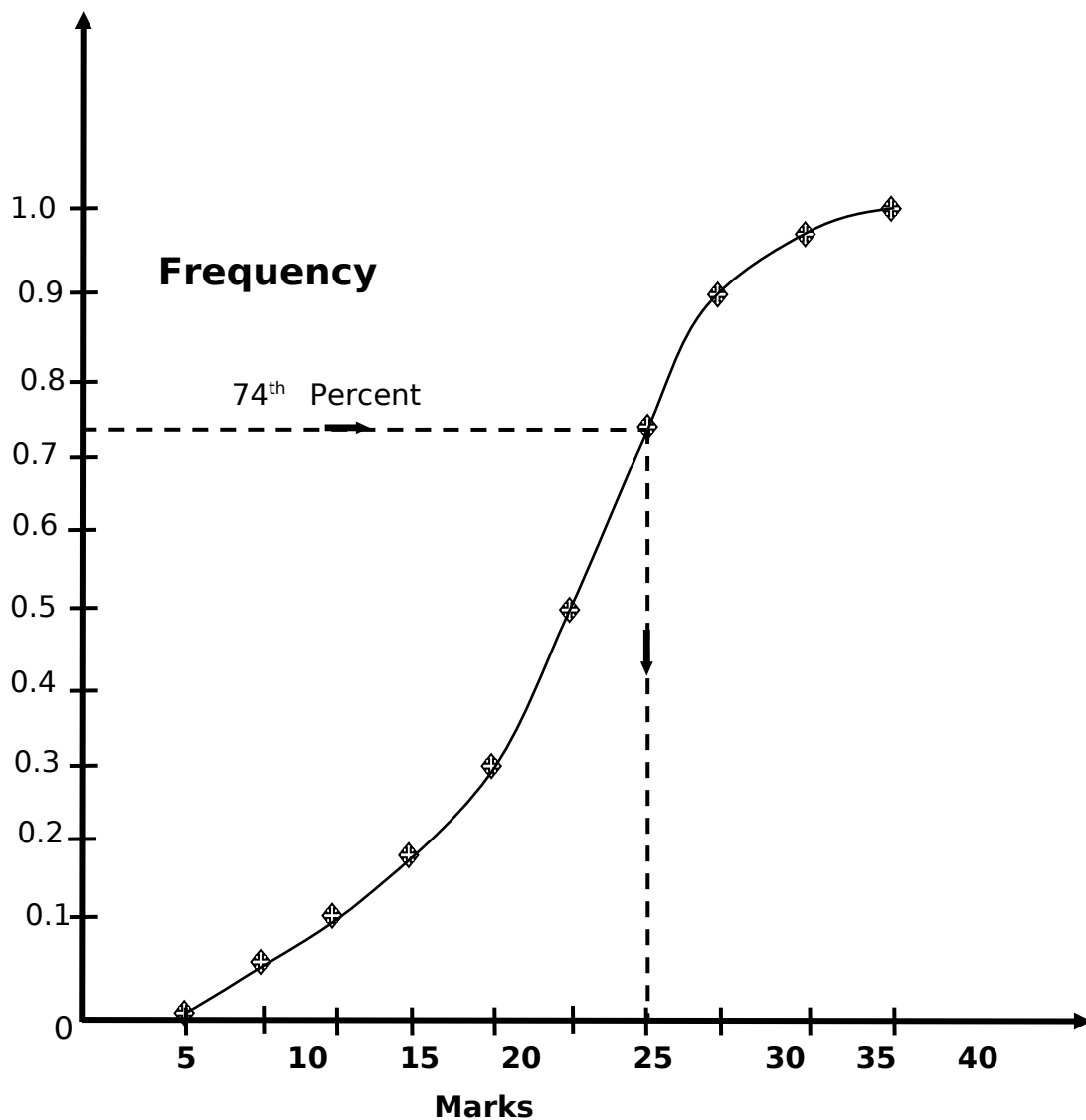
Marks	0 - 4	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49
Frequency	0	4	6	10	14	24	28	19	9	6

- Construct the relative cumulative frequency curve for the above mentioned data.
- From the curve, estimate the 74th percentile.

a) Relative cumulative frequency =
$$\frac{\text{Class interval cumulative frequency}}{\text{Total frequency}}$$

Mark	Frequency	Cumulative Frequency	Relative Cumulative Frequency
0 - 4	0	0	$\frac{0}{120}=0$
5 - 9	4	4	$\frac{4}{120}=0.03$
10 - 14	6	10	$\frac{10}{120}=0.08$
15 - 19	10	20	$\frac{20}{120}=0.17$
20 - 24	14	34	$\frac{34}{120}=0.28$
25 - 29	24	58	$\frac{58}{120}=0.48$
30 - 34	28	86	$\frac{86}{120}=0.72$

35 - 39	19	105	$\frac{105}{120}=0.88$
40 - 44	9	114	$\frac{114}{120}=0.95$
45 - 49	6	120	$\frac{120}{120}=1.00$



b) 74th Percentile

$$= \frac{74}{100} (1.00)$$

$$= 0.74$$

$$= \underline{\underline{34.5 \text{ Marks}}}$$

Variance and Standard Deviation

1.1 Ungrouped Data

Variance

This is a measure of spread that measures the distances or spread of data about the mean. It is found by the formula.

$$\text{Variance} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Where $\bar{x}$ is the mean of $x_1, x_2, x_3, \dots, x_i$ and n is the number of observations.

An alternative method that can be used to calculate the variance for ungrouped data is:

$$\text{Variance} = \frac{\sum x^2}{n} - (\bar{x})^2$$

1.2 Standard Deviation

The standard deviation is the positive square root of the variance.

$$\text{Standard deviation} = \sqrt{\text{Variance}} \\ (\text{S.D})$$

It can be calculated by the formula

$$\text{Standard deviation} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Alternatively, the standard deviation of the ungrouped data can be found by the formula

$$\text{S.D} = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$$

Where $\bar{x}$ is the mean

Question 1

Calculate the variance and standard deviation of the following set of data.

8, 5, 10, 25 and 32.

Solution

$$\text{Mean } \bar{x} = \frac{\sum x}{n}$$

$$\text{Mean } \bar{x} = \frac{8+5+10+25+32}{5}$$

$$\bar{x} = \frac{80}{5}$$

$$\bar{x} = \mathbf{16}$$

Method 1

Using
$$\text{Variance} = \frac{\sum (x - \bar{x})^2}{n}$$

Score x	Deviation $x - \bar{x}$	$(x - \bar{x})^2$
8	-8	64
5	-11	121
10	-6	36
25	9	81
32	16	256
$\sum x = 80$	$\sum (x - \bar{x}) = 0$	$\sum (x - \bar{x})^2 = 558$

2.0. GROUPED DATA

2.1 THE MEAN OF GROUPED DATA

The mean of grouped data also called the estimate of the mean is denoted by $\bar{x}$ is calculated using the formula

$$\text{Mean } (\bar{x}) = \frac{\sum fx}{\sum f}$$

2.2 THE STANDARD DEVIATION OF GROUPED DATA

The formula used to calculate the standard deviation of the grouped data is

$$\text{Standard deviation} = \sqrt{\frac{\sum f(x - \bar{x})^2}{n}}$$

Where $n = \sum f$

$$\bar{x} = \frac{\sum fx}{\sum f}$$

The alternative formula for calculating the standard deviation is

$$\text{Standard deviation} = \sqrt{\frac{\sum fx^2 - \frac{(\sum fx)^2}{\sum f}}{\sum f}}$$

where

$$\bar{x} = \frac{\sum fx}{\sum f}$$

$$\text{Variance} = \frac{\sum (x - \bar{x})^2}{n}$$

$$= \frac{558}{5}$$

$$= 111.6 \text{ (1d.p.)}$$

$$\text{S.D} = \sqrt{\frac{558}{5}}$$

$$= \sqrt{\frac{558}{5}}$$

$$= 10.56 \text{ (2d.p.)}$$

$$\text{Standard deviation} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}}$$

8	64
5	25
10	100
25	625
32	1024
$\sum x = 80$	$\sum x^2 = 1838$

x

$\bar{x}^2$

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{80}{5}$$

$$\bar{x}=16$$

$$\begin{aligned}\text{Standard deviation} &= \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \\ &= \sqrt{\frac{1838 - 16}{5}} \\ &= \sqrt{111.6} \\ &= 10.56\end{aligned}$$

Question

An intelligence quotient test that was taken by pupils at Kitwe Boys Secondary School in Kitwe Showed the following results.

Time taken (Minutes)	$0 < x < 1$	$1 < x < 2$	$2 < x < 3$	$3 < x < 5$	$5 < x < 10$
Frequency	10	15	25	40	25

Find

- (a) The estimate of the mean
(b) The standard deviation

Solution

To calculate the estimate of the mean and standard deviation, we use the mid-interval or midpoint x

Time	x	f	fx	fx^2
$0 \leq x \leq 1$	0.5	10	5	2.5
$1 < x \leq 2$	1.5	15	22.5	33.75
$2 < x \leq 3$	2.5	25	62.5	156.25
$3 < x \leq 5$	4	40	160	640
$5 < x \leq 10$	7.5	25	187.5	1406.25
		$\Sigma f = 115$	$\Sigma fx = 437.5$	$\Sigma fx^2 = 2238.75$

(a) estimate of the mean

$$\bar{x} = \frac{\sum fx}{\sum f}$$
$$\bar{x} = \frac{437.5}{115}$$
$$\bar{x} = 3.80$$

$$S.D = \sqrt{\sum f_x - i(x) i}$$

$$(b) S.D = \frac{\sqrt{\sum f}}{\sum f}$$

$$S.D = \frac{\sqrt{2238.75 - 3.80}}{115}$$

$$S.D = 2.2 (Id.p)$$

$$(c) (i) \text{ Median} = \frac{1}{2} Q$$

$$Q_2 = \frac{1}{2} \times 55$$

$$= 27.5$$

$$Q_2 = \underline{\underline{7.8 \text{ Minutes}}}$$

$$(ii) \text{ Lower Quartile} = \frac{1}{4} Q$$

$$Q_1 = \frac{1}{4} \times 55$$

$$= 13.75$$

$$Q_1 = 5.2 \text{ Minutes}$$

$$(iii) \text{ Upper quartile} = \frac{3}{4} Q$$

$$Q_3 = \frac{3}{4} \times 55$$

$$= 41.25$$

$$= \underline{\underline{9.7 \text{ Minutes}}}$$

$$(iv) \text{ Interquartile range} = \text{Upper quartile} - \text{Lower quartile}$$

$$= Q_3 - Q_1$$

$$9.7 - 5.2$$

$$\underline{\underline{4.5 \text{ Minutes}}}$$

$$(v) \text{ Semi - interquartile range} = 1.2 \times \text{interquartile range}$$

$$\frac{1}{2} \times 4.5 \text{ Minutes}$$

$$2.25 \text{ Minutes}$$

$$(vii) 60^{\text{th}} \text{ Percentile} = 60 \times \frac{56}{100}$$

$$= 33.6$$

$$\underline{\underline{60\% \text{ of } Q = 33.6}}$$

$$\underline{\underline{= 8.45 \text{ Minutes}}}$$

$$(d) 55 - 17$$

$$= 38 \text{ Pupils}$$

$$P(> 6 \text{ Minutes}) = \frac{38}{55}$$

$$= \frac{38}{55}$$

$$(e) (i) P(< 9 \text{ Minutes}) = \frac{37}{55}$$

$$= \frac{37}{55}$$

$$(ii) P(< 12 \text{ Minutes and } < 12 \text{ Minutes})$$

$$P(> 2 \text{ Minutes}) \times P(< 12 \text{ Minutes})$$

$$\frac{5}{55} \times \frac{4}{54} = \frac{2}{297}$$

$$= \frac{2}{297}$$

a) $\text{Mean}(\bar{x}) = \frac{\sum fx}{\sum f}$ $\bar{x} = \frac{2 \times 6 + 5 \times 11 + 8 \times 20 + \dots}{55}$

$\bar{x} = \frac{12 + 55 + 160 + 132 + 70}{55} =$

$\bar{x} = \underline{\underline{7.8 \text{ Minutes (I.d.p)}}}$

Waiting time (mm)	≤ 3	≤ 6	≤ 9	≤ 12	≤ 15
Frequency	16	17	37	50	55

