TEC 106 – PROBABILITY AND STATISTICS

SESSION ONE

Course content

1. Tabular and Graphical representation of Samples
2. Random experiments and events
3. Random variables
4. Probability Distribution

INTRODUCTION TO STATISTICS

STATISTICS – Science that involves the manipulation of the mass of numerical data emanating from activities of interest into forms which useful conclusions can be drawn.

Terms used in Statistics

1. Statistical unit – Unit of reference used in a compiled set of data
2. Population – Collection or set of individual objects of measurements whose properties are to be analyzed
3. Parameter – Numerical characteristics of an entire population
4. Sample – subset of a population
5. Data – Numerical value of the statistical unit associated with one element of a population or a sample

a) *Qualitative/Attribute data* – focuses on quality type of description of the subject. Eg: Colour,

b) *Quantitative/Variable data* – Results from counts or measurements. Can be in two forms: Discrete(countable in whole forms eg: People, cars…) and Continuous(measured on a continuous scale eg: Temperature, Mass…)

6. <u>Random selection</u> – generation of a sample by giving equal change for all to be selected

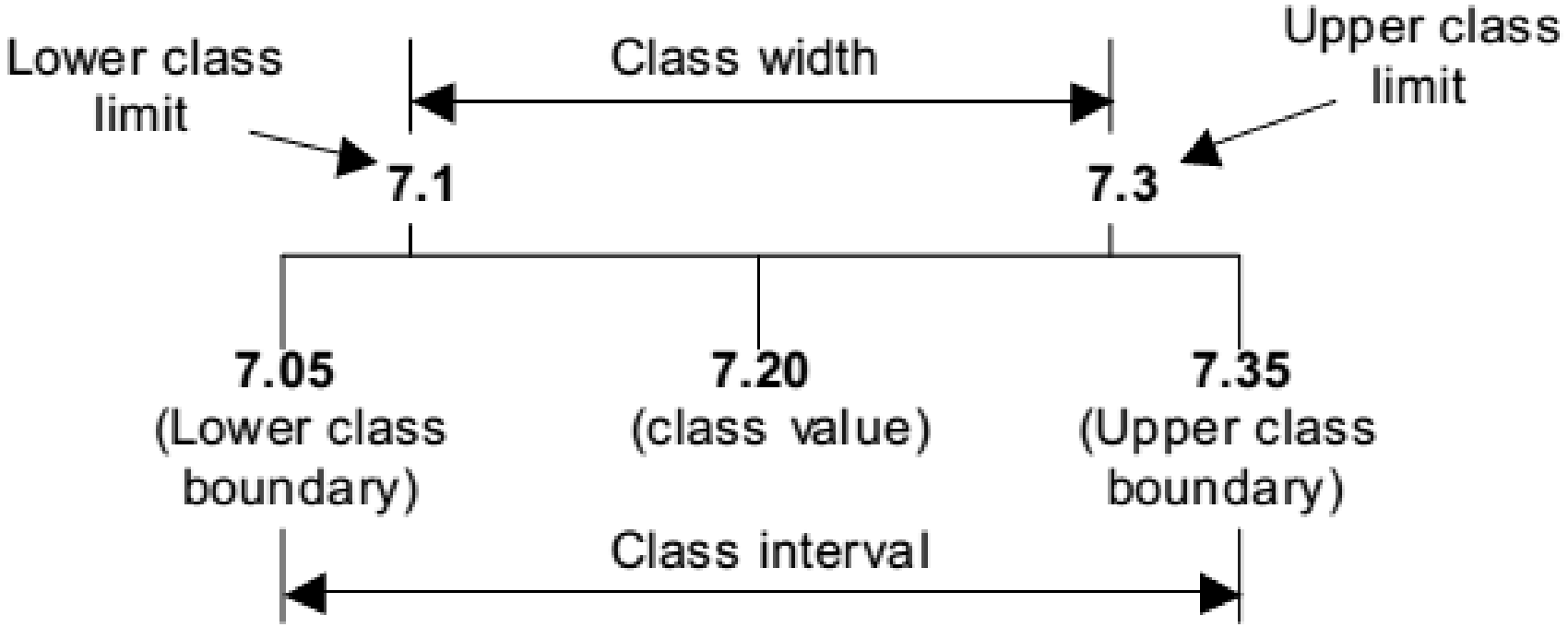
Phases of a Statistical Experiment

a) Formulation of the problem
b) Design of the experiment
c) Collection of data
d) Mathematical description/organization of data
e) Analysis of data
f) Interpretation of data

TABULAR AND GRAPHICAL REPRESENTATION OF SAMPLES

1. Tabular – Frequency distribution table
2. Graphical
   - *Plots of Absolute frequency* (Bar chart, Dot frequency diagram, Cumulative frequency curve or Ogive)
   - *Plots of relative Frequency* (Frequency histogram, frequency polygon, Cumulative frequency function)

Grouping of Data



1. Class interval = Upper class boundary − lower class boundary

2. Class Width =
   Upper class limit − lower class limit

3. Class value =

$$\left(\frac{\text{Upper class limit} + \text{lower class limit}}{2}\right) \text{ or}$$

$$\left(\frac{\text{Upper class boundary} + \text{Lower class boundary}}{2}\right)$$

- $Upper\ class\ boundary = Upper\ class\ limit + \ No_{decimal\ places}/2$

- $Lower\ class\ boundary = Lower\ class\ limit - \ No_{decimal\ places}/2$

Procedure for Grouping data

1. **Determine the range**

$$Range = Largest\ value - lowest\ value$$

2. **Determine the class interval (CI)**

$$\text{CI} = \frac{\text{Range}}{\text{no of desired classes}}$$

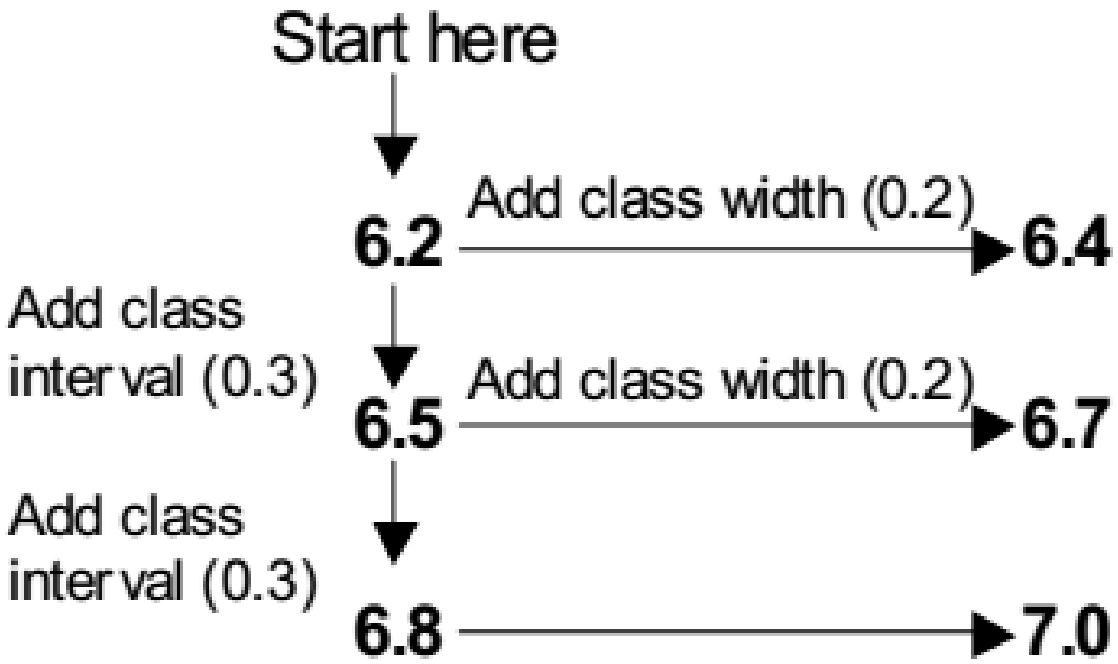
3. **Determine the class width (CW)**

$$\text{CW} = \text{CI} - 10^{-\text{No of dp}}$$

eg: if CI = 0.3, No of dp = 1

$$\text{CW} = 0.3 - 10^{-1} = 0.3 - 0.1 = 0.2$$

4. Construct the classes iteratively



Tabular representations of data

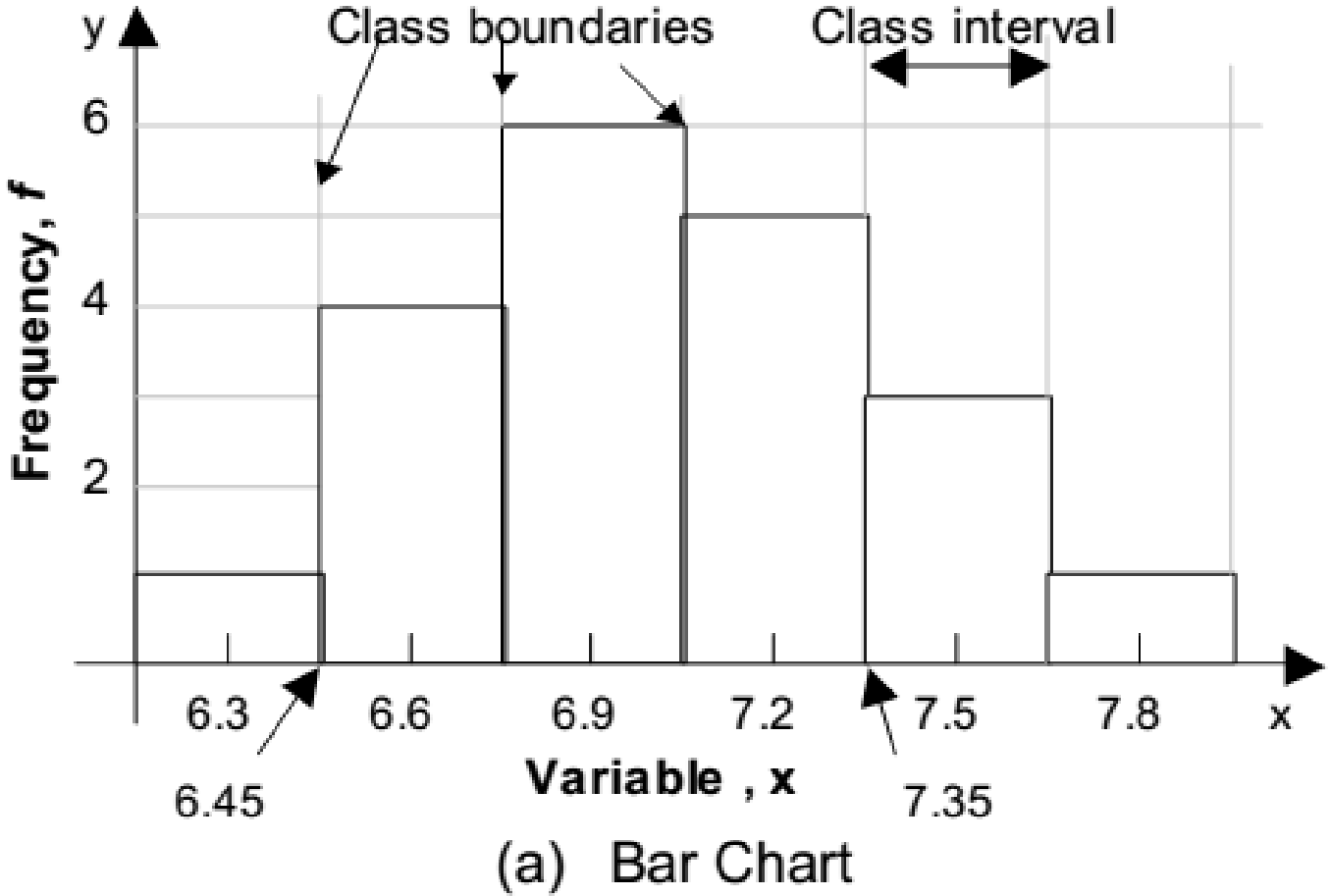
Table 2.1 Different conventions for representing Class intervals

A (metres)	B (metres)	C (metres)	D (metres)
0-	0 and under 5	0 -5	0 – 4
5-	5 " 10	5 - 10	5 – 9
10-	10 " 15	10 - 15	10 – 14
15-	15 " 20	15 - 20	15 -19
20-	20 " 25	20 - 25	20 -29

Variable X [mm]	Class value	(1) f	(2) $r.f$	(3) c.a.f	(4) c.r.f
6.2 - 6.4	6.3	1	0.05	1	0.05
6.5 - 6.7	6.6	4	0.20	5	0.25
6.8 - 7.0	6.9	6	0.30	11	0.55
7.1 - 7.3	7.2	5	0.25	16	0.80
7.4 - 7.6	7.5	3	0.15	19	0.95
7.7 - 7.9	7.8	1	0.05	20	1.00
		n =20	$\sum$ r.f = 1.0		

Graphical plots of data

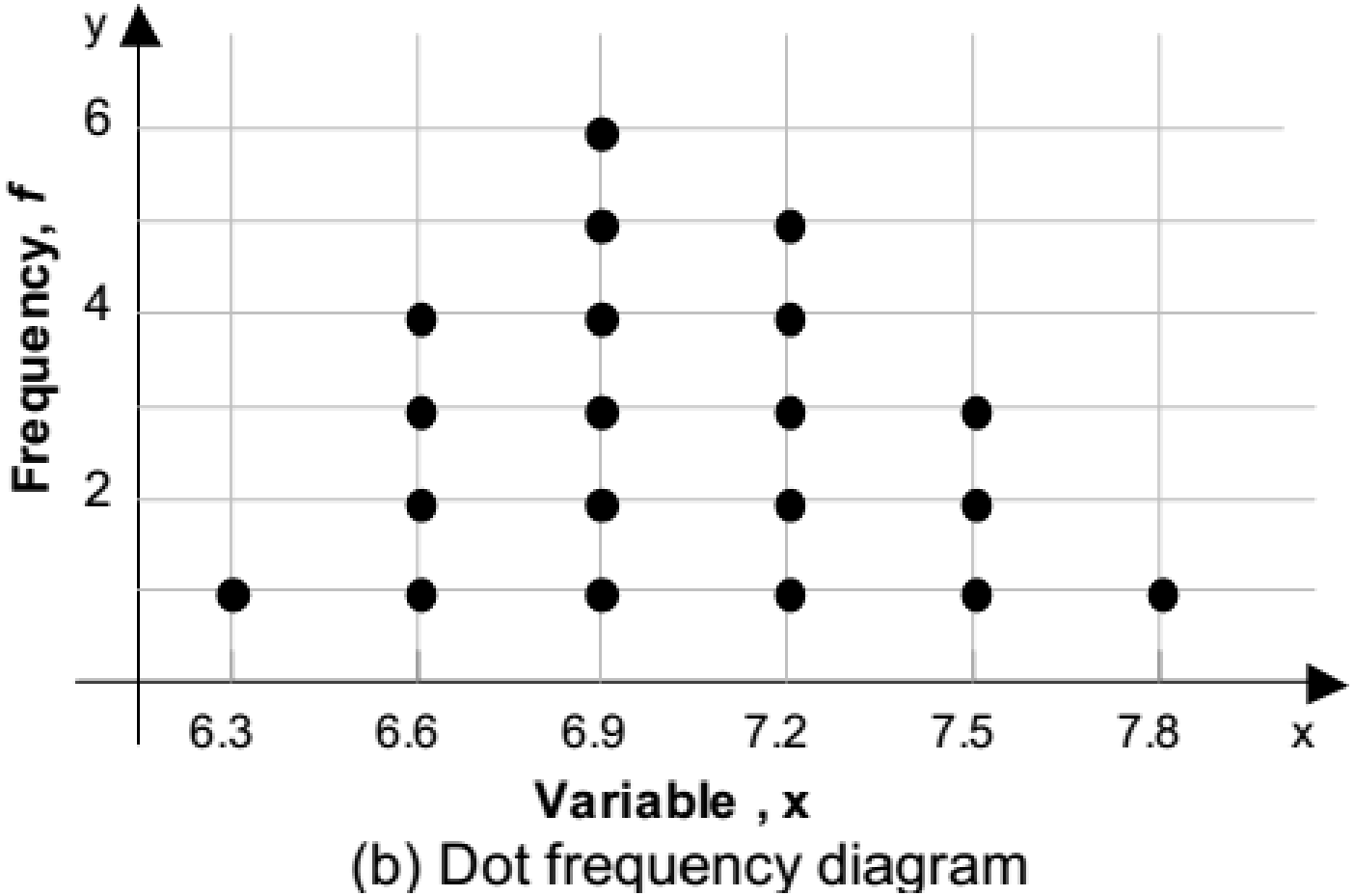
1. BAR CHART
   - Bar plot of Frequency ‘f’ vs Class value ‘x’



(a) Bar Chart

2. DOT FREQUENCY DIAGRAM

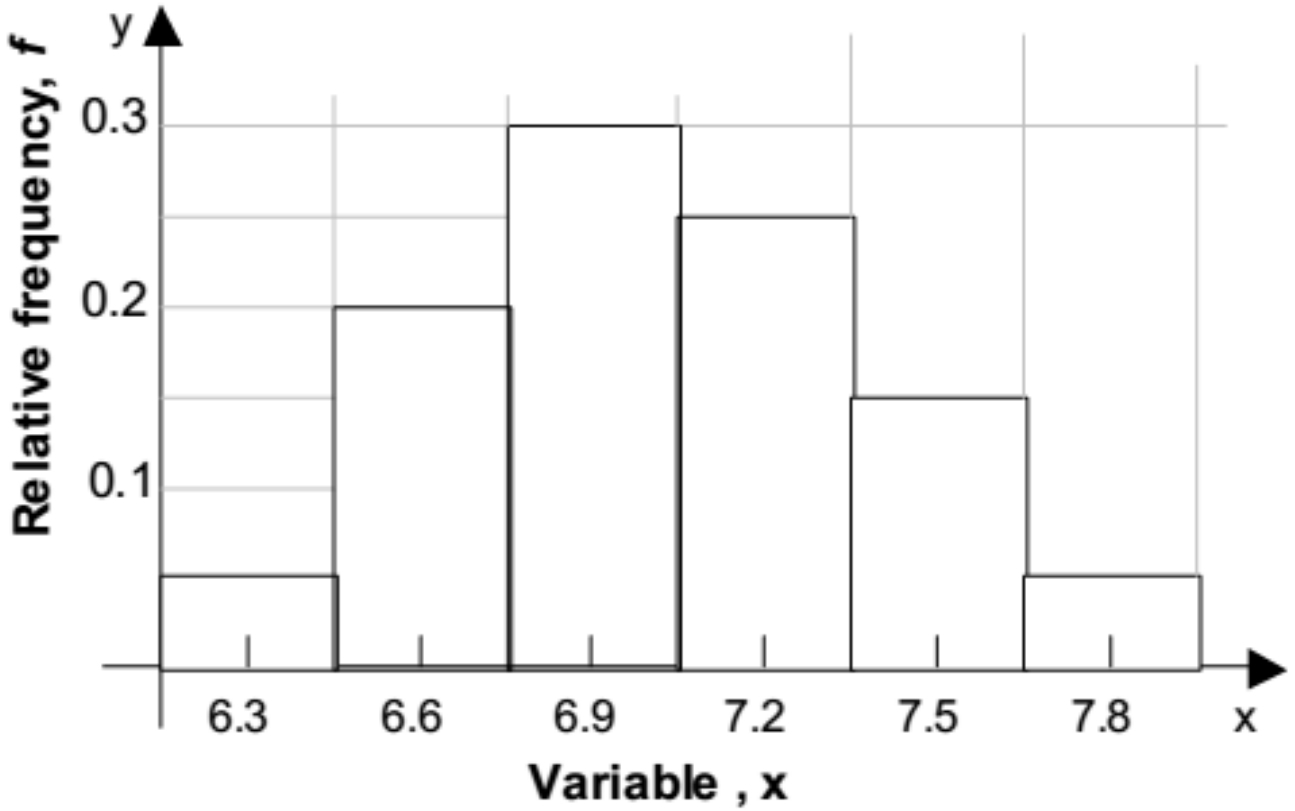
- Dot plots of Frequency ‘f’ vs class values ‘x’



(b) Dot frequency diagram

3. HISTOGRAM

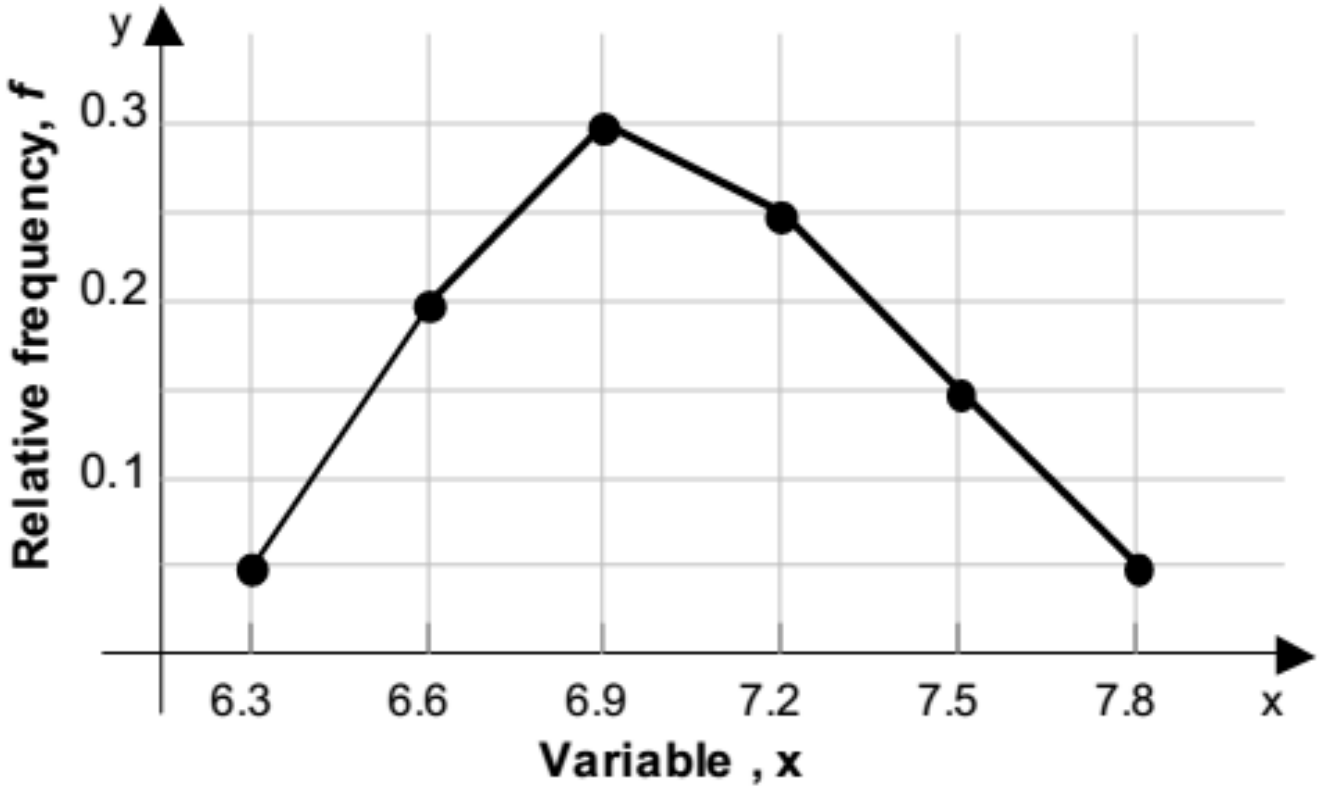
- Bar plot of Relative Frequency ‘f’ vs class values ‘x’



(b) Histogram

4. FREQUENCY POLYGON

- Line plot of Relative Frequency ‘f’ vs class values ‘x’

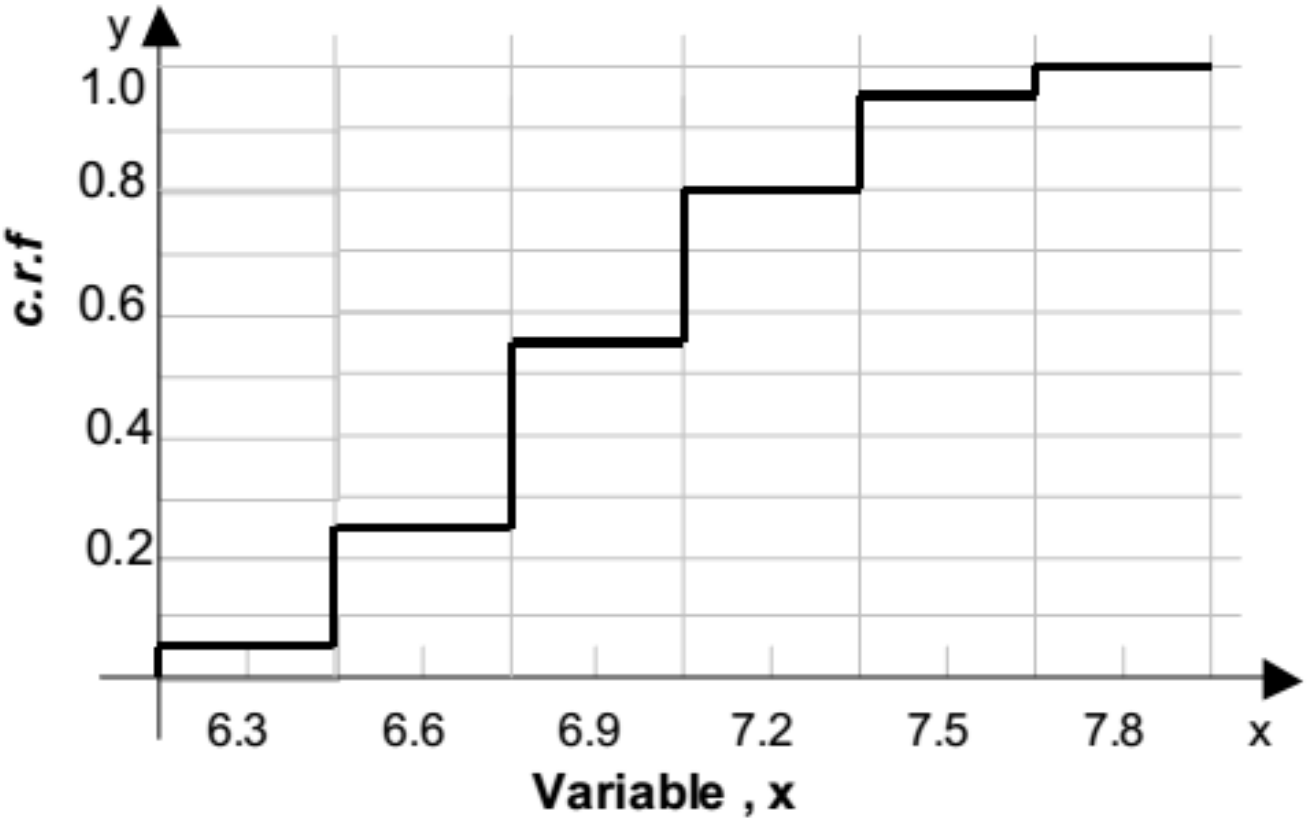


(b) Frequency polygon

Figure 2.3 Histogram and Frequency Polygon

5. CUMULATIVE FREQUENCY FUNCTION

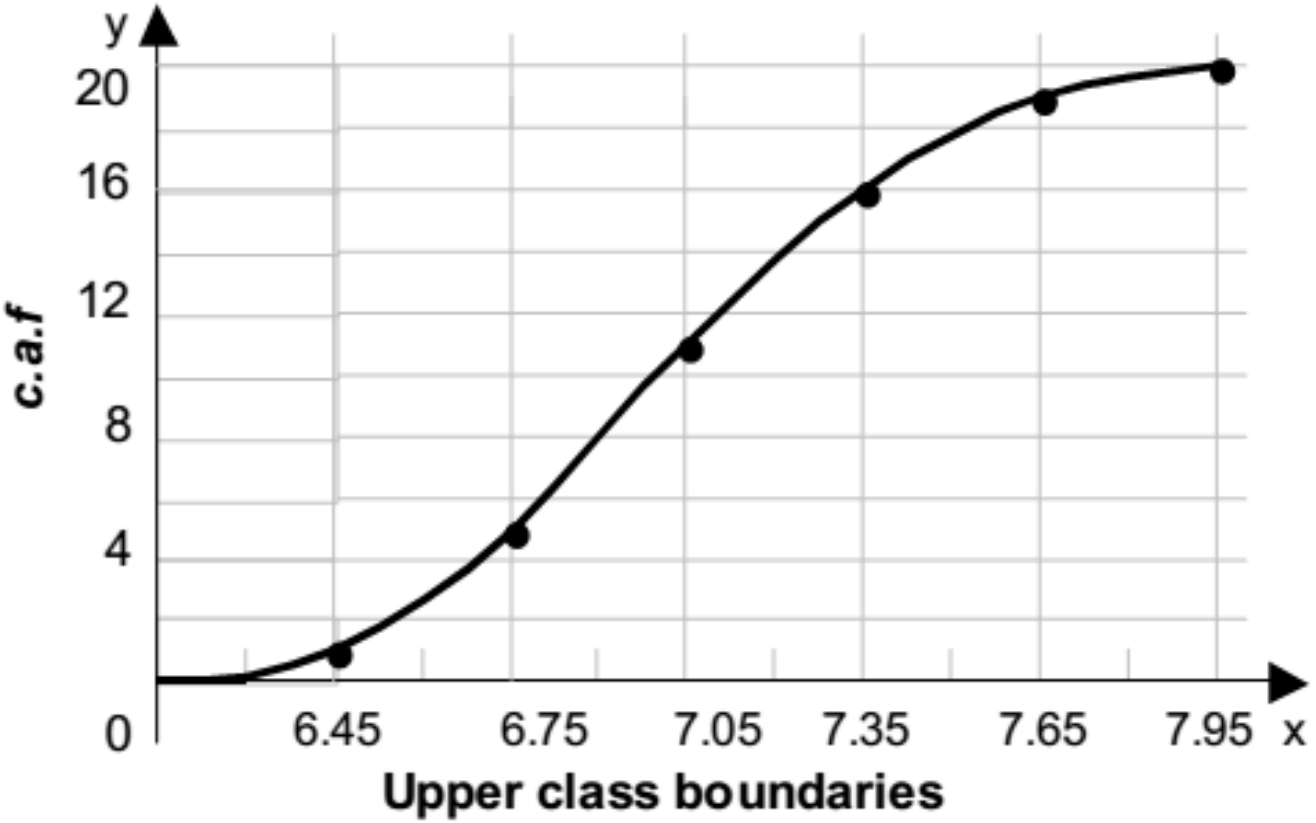
- Line plot of Cumulative relative frequency ‘CRF’ vs class values ‘x’



(a) Cumulative frequency function

6. CUMULATIVE FREQUENCY CURVE (OGIVE)

– Line plot of Cumulative absolute frequency ‘CAF’ vs upper class boundaries ‘UCB’



(b) Cumulative frequency curve (ogive)

7. STEM PLOT

Stems	Leaves
150	1
140	
130	
120	2 6
110	4 5 7 9
100	1 2 2 2 5 7 9 9
90	0 2 3 4 4 5 7 8 9 9
80	1 1 4 7 8
Key: 110	7 represents an IQ score of 117

Eg: 120 | 6

Represents 126