The Quartiles

Case1) Ungrouped data

There are 3 types of quartiles

1) The Lower Quartile, the 1st Quartile or 25 Percentile denoted by Q1
   Position Given by (n+1)/4

2) The Mid Quartile, the 2nd Quartile or the 50 percentile or the Median denoted by Q2
   Position given by (n+1)/2

3) The Upper Quartile, the 3rd Quartile or the 75 percentile denoted by Q3
   Position given by 3(n+1)/4

Find the 1st & 3rd Quartiles, Q1 And Q3,

Of The Following Set Of Numbers

: 347, 242, 146, 391, 249, 567, 277, 218.319

Step 1)

This is ungrouped data. First arrange the data in ascending order.

146,218,242,249,277,319,347,391,567

Now, find median or second quartile.
   Median position: (n+1)/2 = (9+1)/2=10/2=5th position
   Counting median= 277

First quartile (Q1)
   First Quartile Position: (n+1)/4 =10/4=2.5th position
   So average of 2nd and 3rd positions values= (218+242)/2=230

Third quartile (Q3)
   Third Quartile Position: 3(n+1)/4 =30/4=7.5th position
   So average of 7th and 8th position values= (347+391)/2=369.

The Interquartile Range denoted by IQR is given by the difference between the 3rd and the 1st Quartiles.

   IQR= Q3-Q1=369-230=139

Case 2 Grouped data

There are two forms a data may be presented;
   a) Grouped
   b) Ungrouped.
**Grouped and Ungrouped Data**

**Ungrouped Data:-** The data obtained in original form are called raw data or ungrouped data.

**Example 1**
The marks obtained by 25 students in a class in a certain examination are given below:

25, 8, 37, 16, 45, 40, 29, 12, 42, 25, 14, 16, 16, 20, 10, 36, 33, 24, 25, 35, 11, 30, 45, 48.

This is ungrouped data or raw data.

**Array:** - An arrangement of ungrouped data in ascending or descending order of magnitude is called an array.

Arranging the marks of 25 students in ascending order, we get the following array.

8, 10, 11, 12, 14, 16, 16, 16, 20, 24, 25, 25, 25, 29, 30, 33, 35, 36, 37, 40, 40, 42, 45, 45, 48.

**Grouped Data**

To put the data in a more condensed form, we make groups of suitable size, and mention the frequency of each group. Such a table is called a grouped frequency distribution table.

**Class-Interval: for example (20----30)**

Each class is bounded by two figures, which are called class limits. The figure on the left side of a class is called its lower limit (20) and that on its right is called its upper limit (30).

**Types of Grouped Frequency Distribution**

1. **Exclusive Form (or Continuous Interval Form):** A frequency distribution in which the upper limit of each class is excluded and lower limit is included, is called an exclusive form.

**Example:** Suppose the marks obtained by some students in an examination are given. We may consider the classes 0 – 10, 10 – 20 etc. in class 0 – 10, we included 0 and exclude 10. In class 10 – 20, we include 10 and exclude 10 and exclude 20.

(i) **Continuous Interval Form (or Exclusive Form)**

**Example1:** Given below are the marks obtained by 40 students in an examination:

3, 25, 48, 23, 17, 13, 11, 9, 46, 41, 37, 45, 10, 19, 39, 36, 34, 5, 17, 21,
Arrange the data in ascending order and present it as a grouped data in:

(i) Continuous Interval form, taking class-intervals 1 – 10, 10 – 20, etc.

Step 1)

: Arrange the marks in ascending order, we get:

3, 3, 5, 8, 9, 10, 11, 12, 13, 15, 17, 17, 17, 19, 19, 19, 20, 21, 21, 23,

23, 25, 25, 28, 28, 32, 32, 32, 33, 34, 34, 36, 37, 39, 39, 41, 45, 46, 48, 48,

We may now classify them into groups as shown below:

(ii) Continuous Interval Form (or Exclusive Form)

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Tally Marks</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 10</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>10 – 20</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>20 – 30</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>30 – 40</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>40 – 50</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

Here, the class 1 – 10 means, marks obtained from 1 to 9, i.e., excluding 10.

NB) in this type of interval the limits are just the lower and upper values and the class interval is the difference between them e.g. 30---40 class size h=40-30=10

2 Inclusive From (or Discontinuous Interval From):

A frequency distribution, in which each upper limit and well as lower limit is included, is called an inclusive form. Thus, we have classes of the form 0 – 10, 11 – 20, 21 – 30 etc. In 0 – 10, both 0 and 10 are included.

Example1: Given below are the marks obtained by 40 students in an examination:

3, 25, 48, 23, 17, 13, 11, 9, 46, 41, 37, 45, 10, 19, 39, 36, 34, 5, 17, 21,

39, 33, 28, 25, 12, 3, 8, 17, 48, 34, 15, 19, 32, 32, 19, 21, 28, 32, 20, 23,

Solution: Arrange the marks in ascending order, we get:

3, 3, 5, 8, 9, 10, 11, 12, 13, 15, 17, 17, 17, 19, 19, 19, 20, 21, 21, 23,

23, 25, 25, 28, 28, 32, 32, 32, 33, 34, 34, 36, 37, 39, 39, 41, 45, 46, 48, 48,

We may now classify them into groups as shown below:
(i) **Discontinuous Interval Form (or Inclusive Form)**

(i) Discontinuous Interval form, taking class-intervals 1 – 10, 11 – 20, etc.

---

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Tally Marks</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 9</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>10 – 19</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>20 – 29</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>30 – 39</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>40 – 49</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

Note that the class 1 – 10 means, marks obtained from 1 to 10, including both.

**NB** in this type of interval the limits are the lower and upper values we subtract 0.5 from lower value add 0.5 to the upper value. The class interval is the difference between them e.g. 30---39 the limits would be 29.5----39.5. class size h=39.5-29.5=10

---

**MEDIAN**

**Steps for Grouped Data Median:**

1. Locate which class contains the median by using the following formula:
   
   Location of median = \( \frac{15.5}{2} \)  
   
   This formula tells you what “place” the median would be in if the numbers were placed in order from smallest to largest! In this example, the median is located halfway between the 15th and the 16th numbers.

2. Now figure out which class would contain the median. (Remember the median in this example is between the 15th and the 16th numbers in order from smallest to largest.)

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency f</th>
<th>Cumulative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 up to 5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5 up to 10</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td><strong>10 up to 15</strong></td>
<td><strong>12</strong></td>
<td><strong>21</strong></td>
</tr>
<tr>
<td>15 up to 20</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>20 up to 25</td>
<td>3</td>
<td>30</td>
</tr>
</tbody>
</table>

The median is somewhere between 10 and 15. We know this because we can look at the *Cumulative Frequency* and see that the 15th and the 16th numbers are in this class! (In fact, the 10th through the 21st numbers were in this class)

3. Do the Math!

**This is one of the trickier formulas for this exam! Please take close note of**
what the following variables stand for in this particular problem:

$$\text{Median} = L + \frac{(n/2 - CF)}{f} \cdot h$$

$L = \text{Lower limit of the class that contains the median} = 10$
$n = \text{Number of numbers} = 30$
$CF = \text{Number of numbers before the class containing the median} = 9$
$f = \text{number of numbers in the class containing the median} = 12$
$i = \text{class interval (size)} = \text{upper Limit} - \text{Lower limit}$

$$\text{Median} = 10 + \frac{(30/2 - 9) \cdot 10}{12}$$

**MODE (GROUPED) = MOST**
The mode for a grouped data problem is the midpoint of the class with the highest frequency ($f$).
The class 10 – 15 has the highest frequency (12), so the mode is the midpoint of this class. **MODE = 12.5**

**Please take close note that the formulas used for computing the mean, the median, and the standard deviation for GROUPED data, all contain lowercase “$f$” in them!!!! I would recommend that you use this fact to help remember which formulas are appropriate for the grouped data problems.**

**RANGE (GROUPED)**
Range (Grouped) = Upper Limit of Highest Class – Lower Limit of Lowest Class
In our example, 0-5 is the lowest class and 0 is the lower limit of this class. Also, 20-25 is the highest class, and 25 is the upper limit of this class.
Range(Grouped) = 25 – 0 = 25

**Quartiles for Grouped data**

$Q1$ and $Q3$ for data grouped by cell.

$$Q1 = L + \frac{(n/4 - CF) \cdot h}{f} \quad \text{and} \quad Q3 = L + \frac{(3n/4 - CF) \cdot h}{f}$$

Where $L = \text{the lower limit of the cell which contains the respective quartile}$, $F$ is the sum of the frequencies below that cell, and $f_{Q1}$ and $f_{Q3}$ are respectively the frequencies of the cells which contain the first and third quartiles.

. **Inter quartile Range: IQR = Q3 - Q1**

. **Percentiles and deciles** calculated in a similar manner.

For raw data the $k$th percentile
The $k$th deciles
For grouped data
Where are defined analogously to the way they are defined in the formulas for quartiles
16. **Range** = Maximum - Minimum
17. **Mode** = the most frequently occurring value. In a grouped data it is the midpoint of the modal class.

(b) The following table gives the weekly wages earned by all employees in a factory.

<table>
<thead>
<tr>
<th>Weekly Wage</th>
<th>No. of Employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>150 – 169</td>
<td>16</td>
</tr>
<tr>
<td>170 – 189</td>
<td>31</td>
</tr>
<tr>
<td>190 – 209</td>
<td>22</td>
</tr>
<tr>
<td>210 – 229</td>
<td>14</td>
</tr>
<tr>
<td>230 – 249</td>
<td>10</td>
</tr>
<tr>
<td>250 – 269</td>
<td>7</td>
</tr>
</tbody>
</table>

Using the above data:
(i) Present the data in a cumulative frequency distribution table.
(2 marks)
(ii) Draw a Histogram and Ogive to represent the above data.
(4 marks)
(iii) Calculate the mean and the median.
(6 marks)
(iv) Calculate the inter-quartile range.
(4 marks)