CHAPTER 7- STATISTICS

7.1 MEASURE OF CENTRAL TENDENCY

7.1.1 Ungrouped Data
We have learned how to find mode, median and mean of ungrouped data in Form Three. So this is just a revision for us to remind on it.

Mean
The formula is $\bar{x} = \frac{\sum x}{N}$ where

(i) $\bar{x}$ represents the mean of the data.
(ii) $\sum x$ represents the summation or total of data
(iii) $N$ represents the number of data

Example:
Find the mean of the following set of data.
73, 76, 70, 65, 81, 79.

Solution:
Let $x$ represent a value in the set of data.
$\sum x = 73 + 76 + 70 + 65 + 81 + 79$
$= 444$
$N = 6$
The formula is $\bar{x} = \frac{\sum x}{N}$, then

$$\bar{x} = \frac{444}{6} = 74$$

Mode
Mode of a set of data is the value that has the highest frequency.

Example:
Given set of data which is 1, 8, 6, 4, 6, 3, 6, 5. Determine the mode.

Solution:
Mode is 6.
Median
Median is the value that lies in the middle of a set of data.

Example:
Given set of data which is 1, 8, 6, 4, 6, 3, 6, 5, 7. Determine the median.

Solution:
At first, arrange the numbers in order.
1, 3, 4, 5, 6, 6, 6, 7, 8.

4 values
4 value
Median

Hence, the median is 6.

7.1.2 Grouped Data

Modal Class
The class having the highest frequency is known as the modal class.

Example:
The following frequency distribution table shows the mass of 45 students in a class.

<table>
<thead>
<tr>
<th>Mass(kg)</th>
<th>50-54</th>
<th>55-59</th>
<th>60-64</th>
<th>65-69</th>
<th>70-74</th>
<th>75-79</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>4</td>
<td>12</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Determine the modal class.

Solution:
The class having the highest frequency is 60 – 64 kg. Hence, the modal class is 60 – 64 kg.

Finding mode from a histogram

Example:
1. Determine the modal class in the histogram which is the class 60-69.
2. Join the top vertices of the bar representing the modal class to the vertices if the adjacent bars, as shown in the diagram.
3. Determine the value on the horizontal axis at the intersection of the two lines. This value obtained represents the mode. Hence, the mode of the distribution is 63.5 kg.
Mean

The formula is \( \bar{x} = \frac{\sum fx}{\sum f} \) where

(i) \( \sum fx \) represents the summation or total of the frequency and the class mark that is the mid-point of the class.

(ii) \( \sum f \) represents the summation or total of the frequency.

Example:

The following frequency distribution table shows the heights of 40 students in a class.

<table>
<thead>
<tr>
<th>Height(cm)</th>
<th>140-149</th>
<th>150-159</th>
<th>150-169</th>
<th>170-179</th>
<th>180-189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>1</td>
<td>11</td>
<td>18</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

Calculate the mean height of the students.

Solution:

<table>
<thead>
<tr>
<th>Height(cm)</th>
<th>140-149</th>
<th>150-159</th>
<th>150-169</th>
<th>170-179</th>
<th>180-189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency(f)</td>
<td>1</td>
<td>11</td>
<td>18</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Midpoint(x)</td>
<td>144.5</td>
<td>154.5</td>
<td>164.5</td>
<td>174.5</td>
<td>184.5</td>
</tr>
</tbody>
</table>

The formula is \( \bar{x} = \frac{\sum fx}{\sum f} \).

\[
\bar{x} = \frac{1(144.5) + 11(154.5) + 18(164.5) + 8(174.5) + 2(184.5)}{1 + 11 + 18 + 8 + 2}
\]

\[
\bar{x} = \frac{6570}{40} = 164.25 \text{cm}
\]

Calculating the median from the cumulative frequency

Cumulative frequency is the total or summation of frequency.

The formula is \( m = L + \left( \frac{N - F}{f_m} \right) C \) where,

(i) \( L \) = is the lower boundary of the class in which the median lies.

(ii) \( N \) = total frequency

(iii) \( F \) = the cumulative frequency before the class in which the median lies

(iv) \( C \) = the size of the class which is the upper boundary – lower boundary.

(v) \( f_m \) = the frequency of the class in which the median lies.
Example:
The following frequency distribution table shows the life span of 100 light bulbs produced by a factory.

<table>
<thead>
<tr>
<th>Life span(days)</th>
<th>50-54</th>
<th>55-59</th>
<th>60-64</th>
<th>65-69</th>
<th>70-74</th>
<th>75-79</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>6</td>
<td>12</td>
<td>18</td>
<td>24</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

Calculate the median.

Solution

The formula is

\[ m = L + \left( \frac{N - F}{f_m} \right)C \]

\[ m = 64.5 + \left( \frac{100 - 36}{24} \right)(5) \]

\[ = 67.42 \text{ days} \]

**Estimating the median of grouped data from an ogive**

The median is 67.5 days.
7.1.3 Determining the effects on mode, median and mean when some values in a set of data are changed

When every data in a set of data is multiplied or divided by a constant \( c \) and then added or subtracted by a constant \( k \), then the new mean, median and mode are given by:

\[
\begin{align*}
\text{New mean} &= c(\text{original mean}) + k \\
\text{New median} &= c(\text{original median}) + k \\
\text{New mode} &= c(\text{original mode}) + k \\
\end{align*}
\]

Where \( c \) and \( k \) are constants.

Example:
The mean, mode and median of a set of data are 7.5, 8 and 6 respectively. Find the new mean, mode and median if every value of the set of data is
(a) divided by 2
(b) subtracted by 3

Solution:
(a) \( c \) is \( \frac{1}{2} \) and \( k \) is 0

New mean = \( \frac{1}{2} \)(7.5) + 0
= 3.75

New mode = \( \frac{1}{2} \)(8) + 0
= 4

New median = \( \frac{1}{2} \)(6) + 0
= 3

(b) \( c \) is 1 and \( k \) is -3

New mean = 1(7.5) - 3
= 4.5

New mode = 1(8) - 3
= 5

New median = 1(6) - 3
= 3
EXERCISE 7.1
1. Determine the mode of the following sets of data.
   (a) 5, 4, 8, 4, 6, 3
   (b) 2, 8, 10, 9, 8, 7, 8
2. Find the mean of each of the following sets of data.
   (a) 1, 2, 3, 3, 4, 8
   (b) 3, 3, 5, 6, 7, 7, 11, 12, 14
3. State the modal class for the following frequency distribution table.

<table>
<thead>
<tr>
<th>Score</th>
<th>1-3</th>
<th>4-6</th>
<th>7-9</th>
<th>10-12</th>
<th>13-15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

4.

<table>
<thead>
<tr>
<th>Length(mm)</th>
<th>160-169</th>
<th>170-179</th>
<th>180-189</th>
<th>190-199</th>
<th>200-209</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>15</td>
<td>21</td>
<td>27</td>
<td>18</td>
<td>9</td>
</tr>
</tbody>
</table>

By using a scale of 2cm to 10mm on the x-axis and 2 cm to 5 units on the y-axis, draw a histogram and hence, determine the value of mode.

5. Find the value of mean from the following frequency distribution table.

<table>
<thead>
<tr>
<th>Length(mm)</th>
<th>60-69</th>
<th>70-79</th>
<th>80-89</th>
<th>90-99</th>
<th>100-109</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>8</td>
<td>14</td>
<td>19</td>
<td>13</td>
<td>6</td>
</tr>
</tbody>
</table>

6. Find the value of median from the following frequency distribution table.

<table>
<thead>
<tr>
<th>Length(cm)</th>
<th>30-34</th>
<th>35-39</th>
<th>40-44</th>
<th>45-49</th>
<th>50-54</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>14</td>
<td>19</td>
<td>28</td>
<td>23</td>
<td>16</td>
</tr>
</tbody>
</table>

7. Given the integers 4, 4, 8, 9, 13, 15 and 16, find the values of mean, mode and median. Hence, state the new mean, mode and median when each of the integer is
   (a) increased by 4
   (b) reduced by 2
   (c) multiplied by 3
   (d) divided by 5

8. Find the value of mode from the following histogram.
7.2 MEASURE OF DISPERSION

7.2.1 Range of ungrouped data

Range of ungrouped data = largest value – lowest value

7.2.2 Interquartile range of ungrouped data

Given a set of data 2, 4, 6, 7, 10, 12, 16, 18, 19, 20, 21.

\[
\begin{align*}
2, & \quad 4, \quad 6, \quad 7, \quad 10, \quad 12, \quad 16, \quad 18, \quad 19, \quad 20, \quad 21
\end{align*}
\]

First quartile, \( Q_1 \)  
Second quartile, \( Q_2 \) or median  
Third quartile, \( Q_3 \)

Interquartile range = \( Q_3 - Q_1 \)

Interquartile range = 19 – 7  
= 12

\[
\begin{align*}
Q_1 &= \frac{1}{4} \text{ th value of the data} \\
Q_3 &= \frac{3}{4} \text{ th value of the data}
\end{align*}
\]

Finding the range of grouped data

Range of grouped data = largest class mark – lowest class mark

Class mark is the mid-point of the class.

Example:
The following frequency distribution table shows the life span of 100 light bulbs produced by a factory.

<table>
<thead>
<tr>
<th>Life span (days)</th>
<th>50-54</th>
<th>55-59</th>
<th>60-64</th>
<th>65-59</th>
<th>70-74</th>
<th>75-79</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>6</td>
<td>12</td>
<td>18</td>
<td>24</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

Find the range of the distribution.
Solution:

Largest class mark = \( \frac{75 + 79}{2} \)

\[ = 77 \]

Largest class mark = \( \frac{50 + 54}{2} \)

\[ = 52 \]

Range = 77 – 52

\[ = 25 \text{ days.} \]

Finding the interquartile range of grouped data from the cumulative frequency table

Example:
The following frequency distribution table shows the daily wages of 88 workers.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number or workers</td>
<td>10</td>
<td>11</td>
<td>18</td>
<td>20</td>
<td>15</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

Determine the interquartile range.

Solution:

\[ Q_1 = \frac{1}{4}\text{th value of the data} \]

\[ Q_3 = \frac{3}{4}\text{th value of the data} \]

\[ Q_1 = \frac{1}{4} \times 88 \]

\[ = 22 \text{ th data} \]

\[ Q_3 = \frac{3}{4} \times 88 \]

\[ = 66 \text{ th data} \]
The formula to find the first quartile is

\[ Q_1 = L_{Q1} + \left( \frac{1}{4}N - F_{Q1} \right) \frac{C}{f_{Q1}} \]

While the formula to find the third quartile is

\[ Q_3 = L_{Q3} + \left( \frac{3}{4}N - F_{Q3} \right) \frac{C}{f_{Q3}} \]

The difference between the two formulae is \( Q_1 = \frac{1}{4} \) th value of the data while \( Q_3 = \frac{3}{4} \) th value of the data.

\[ Q_1 = 16.5 + \left( \frac{1}{4}(88) - 21 \right) \frac{3}{18} \]

\[ Q_1 = 16.67 \]

\[ Q_3 = 22.5 + \left( \frac{3}{4}(88) - 59 \right) \frac{3}{15} \]

\[ Q_3 = 23.90 \]

The interquartile range is \( Q_3 - Q_1 \),

\[ = 23.90 - 16.67 \]

\[ = RM 7.23 \]
Determining interquartile range of grouped data from an ogive

$Q_1 = \left(\frac{1}{4}\times 40\right)$ th value
$= 10^{th}$ value
$= 189.5$

$Q_3 = \left(\frac{3}{4}\times 40\right)$ th value
$= 30^{th}$ value
$= 259.5$

The interquartile range is $Q_3 - Q_1$, 
$= 259.5 - 189.5$
$= 70$ words per minute.

Determining the varience

1. The formula to find varience of ungrouped data is

$$\sigma^2 = \frac{\sum x^2}{N} - \bar{x}^2$$

2. The formula to find varience of grouped data is

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2$$

where,

$f = \text{frequency of each class}$
$x = \text{class mark}$
$\bar{x} = \frac{\sum fx}{\sum f}$

Example 1:
Determine the varience of set of data 9, 10, 8, 1, 4, 7, 5, 12.

Solution:

$$\bar{x} = \frac{\sum x}{N}$$
$= \frac{9 + 10 + 8 + 1 + 4 + 7 + 5 + 12}{8}$
$= 7$
\[ \sigma^2 = \frac{\sum (x - \bar{x})^2}{N} \]

\[ \sigma^2 = \frac{(9 - 7)^2 + (10 - 7)^2 + (8 - 7)^2 + (1 - 7)^2 + (4 - 7)^2 + (7 - 7)^2 + (5 - 7)^2 + (12 - 7)^2}{8} \]

\[ \sigma^2 = 11 \]

**Example 2:**

The following table shows the number of books published in a year by 30 publishers.

<table>
<thead>
<tr>
<th>Number of books</th>
<th>50-99</th>
<th>100-149</th>
<th>150-199</th>
<th>200-249</th>
<th>250-299</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>4</td>
<td>8</td>
<td>11</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Find the variance of the number of books published.

**Solution:**

Let \( x \) represent the number of books published.

<table>
<thead>
<tr>
<th>Number of books</th>
<th>Class mark, ( x )</th>
<th>( f )</th>
<th>( fx )</th>
<th>( x^2 )</th>
<th>( fx^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-99</td>
<td>74.5</td>
<td>4</td>
<td>298</td>
<td>5550.25</td>
<td>22201</td>
</tr>
<tr>
<td>100-149</td>
<td>124.5</td>
<td>8</td>
<td>996</td>
<td>15500.25</td>
<td>124002</td>
</tr>
<tr>
<td>150-199</td>
<td>174.5</td>
<td>11</td>
<td>1919.5</td>
<td>30450.25</td>
<td>334952.75</td>
</tr>
<tr>
<td>200-249</td>
<td>224.5</td>
<td>5</td>
<td>1122.5</td>
<td>50400.25</td>
<td>252001.25</td>
</tr>
<tr>
<td>250-299</td>
<td>274.5</td>
<td>2</td>
<td>549</td>
<td>75350.25</td>
<td>250700.5</td>
</tr>
</tbody>
</table>

\[ \sum f = 30 \]

\[ \sum fx = 4885 \]

\[ \sum fx^2 = 883857.50 \]

\[ \bar{x} = \frac{\sum fx}{\sum f} \]

\[ \bar{x} = \frac{4885}{30} \]

\[ \bar{x} = 162.83 \]

\[ \sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2 \]

\[ \sigma^2 = \frac{883857.50}{30} - (162.83)^2 \]

\[ \sigma^2 = 2948.3 \]
Determining standard deviation

Standard deviation, \( \sigma = \sqrt{\text{variance}} \)

\[ = \sqrt{\frac{\sum x^2}{N} - x^2} \]

For grouped data,

Standard deviation, \( \sigma = \sqrt{\frac{\sum fx^2}{\sum f} - x^2} \)

**Example 1:**
Determine standard deviation of set of data 9, 10, 8, 1, 4, 7, 5, 12.

**Solution:**
\[
\overline{x} = \frac{\sum x}{N} = \frac{9 + 10 + 8 + 1 + 4 + 7 + 5 + 12}{8} = 7
\]
\[
\sigma^2 = \frac{\sum (x - \overline{x})^2}{N} = \frac{(9 - 7)^2 + (10 - 7)^2 + (8 - 7)^2 + (1 - 7)^2 + (4 - 7)^2 + (7 - 7)^2 + (5 - 7)^2 + (12 - 7)^2}{8}
\]
\[
= 11
\]
\( \sigma = 3.3166 \)

**Example 2:**

The following table shows the number of books published in a year by 30 publishers.

<table>
<thead>
<tr>
<th>Number of books</th>
<th>50-99</th>
<th>100-149</th>
<th>150-199</th>
<th>200-249</th>
<th>250-599</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>4</td>
<td>8</td>
<td>11</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Find the standard deviation of the number of books published.
Solution:

Let \( x \) represent the number of books published.

<table>
<thead>
<tr>
<th>Number of books</th>
<th>Class mark, ( x )</th>
<th>( f )</th>
<th>( fx )</th>
<th>( x^2 )</th>
<th>( fx^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-99</td>
<td>74.5</td>
<td>4</td>
<td>298</td>
<td>5550.25</td>
<td>22201</td>
</tr>
<tr>
<td>100-149</td>
<td>124.5</td>
<td>8</td>
<td>996</td>
<td>15500.25</td>
<td>124002</td>
</tr>
<tr>
<td>150-199</td>
<td>174.5</td>
<td>11</td>
<td>1919.5</td>
<td>30450.25</td>
<td>334952.75</td>
</tr>
<tr>
<td>200-249</td>
<td>224.5</td>
<td>5</td>
<td>1122.5</td>
<td>50400.25</td>
<td>252001.25</td>
</tr>
<tr>
<td>250-299</td>
<td>274.5</td>
<td>2</td>
<td>549</td>
<td>75350.25</td>
<td>250700.5</td>
</tr>
<tr>
<td></td>
<td>( \sum f = 30 )</td>
<td></td>
<td>( \sum fx = 4885 )</td>
<td>( \sum fx^2 = 883857.50 )</td>
<td></td>
</tr>
</tbody>
</table>

\[
\bar{x} = \frac{\sum fx}{\sum f} = \frac{4885}{30} = 162.83
\]

\[
\sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2 = \frac{883857.50}{30} - (162.83)^2 = 2948.3
\]

\[
\sigma = \sqrt{2948.3} = 54.2983
\]

**Determining the effects on measures of dispersion when some values in a set of data are changed**

1. Generally when each data in a set of data is added or subtracted with a constant value, the measures of dispersion of the new set of data do no change.
2. But, if each value in set of data is multiplied with a constant, \( k \) so

   - The new range = \( k \times \) original range
   - The new interquartile range = \( k \times \) original interquartile range
   - The new standard deviation = \( k \times \) original standard deviation
   - The new variance = \( k^2 \times \) original variance
3. If each value in set of data is divided with a constant, \( k \) so

- The new range = original range ÷ \( k \)
- The new interquartile range = original interquartile range ÷ \( k \)
- The new standard deviation = original standard deviation ÷ \( k \)
- The new variance = original variance ÷ \( k^2 \)

**Example:**
The interquartile range and standard deviation of asset of data are 5 and 2.5. Find the new interquartile range and standard deviation if every value of data is divided by 2 followed by an addition of 10.

**Solution:**
New interquartile range = 5 ÷ 2
= 2.5

New standard deviation = 2.5 ÷ 2
= 1.25

**EXERCISE 7.2**
1. Find the value of variance of each of the following sets of integers.
   (a) 5, 6, 6, 10, 12                      (b) 2, 5, 8, 8, 10, 11, 15
2. Find the value of the variance of the following frequency distribution table.

<table>
<thead>
<tr>
<th>Score</th>
<th>0-9</th>
<th>10-19</th>
<th>20-29</th>
<th>30-39</th>
<th>40-49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>5</td>
<td>9</td>
<td>17</td>
<td>12</td>
<td>7</td>
</tr>
</tbody>
</table>
3. Find the standard deviation of the following sets of integers.
   (a) 2, 5, 5, 6, 7                      (b) 1, 5, 6, 8, 10, 11
4. Find the standard deviation of the data in the following frequency distribution table.

<table>
<thead>
<tr>
<th>Score</th>
<th>10-14</th>
<th>15-19</th>
<th>20-24</th>
<th>25-29</th>
<th>30-34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>3</td>
<td>9</td>
<td>12</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>
5. Estimate the interquartile range from the following ogive.
CHAPTER REVIEW EXERCISE

1. Find the value of mean and variance of the set of data of integers 3, 5, 6, 9, 10, 12, 13.

2. Table below shows the masses of 100 children in a kindergarten.

<table>
<thead>
<tr>
<th>Mass(kg)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-29</td>
<td>15</td>
</tr>
<tr>
<td>30-39</td>
<td>25</td>
</tr>
<tr>
<td>40-49</td>
<td>37</td>
</tr>
<tr>
<td>50-59</td>
<td>16</td>
</tr>
<tr>
<td>60-69</td>
<td>4</td>
</tr>
<tr>
<td>70-79</td>
<td>3</td>
</tr>
</tbody>
</table>

(a) Without using an ogive, estimate the median mass.
(b) Calculate the mean and standard deviation of the given distribution.
(c) Draw a histogram and use it to estimate the modal mass.

3. Table below shows the times taken by 100 children to run distance of 50 metres.

<table>
<thead>
<tr>
<th>Time(sec)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-19</td>
<td>10</td>
</tr>
<tr>
<td>20-29</td>
<td>18</td>
</tr>
<tr>
<td>30-39</td>
<td>21</td>
</tr>
<tr>
<td>40-49</td>
<td>23</td>
</tr>
<tr>
<td>50-59</td>
<td>18</td>
</tr>
<tr>
<td>60-69</td>
<td>10</td>
</tr>
</tbody>
</table>

(a) State the modal class of the distribution.
(b) Calculate the mean of the distribution.
(c) Find the interquartile range.

4. The sum $\sum x$ and $\sum x^2$ of ten values and 438 respectively. If two new numbers, 5 and 8, are taken away from the ten values, find the new mean and standard deviation.

5. The mean and standard deviation of a set integers 2, 4, 8, $p$, $q$ are 5 and 2 respectively.
   (a) Find the values of $p$ and of $q$.
   (b) State the mean and standard deviation of set of integers 5, 7, 11, $p+3$, $q+3$

6. The mean and varience of six numbers are 6 and 8 respectively. When a new number $p$ is added, the mean remains the same. Find the values of $p$ and the new varience.

7. The mean of five integers, 3, 4, $m$, 2$m$ and 22 is 10. When 2 is added to each of the five integers, the new median is $\alpha$. Find the value of $m$ and of $\alpha$. 