MAT2001-Statistics for Engineers:Module-1

Dr. Nalliah M

Assistant Professor
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology
Vellore,Tamil Nadu
India.

nalliah.moviri@vit.ac.in

July 20, 2020

Contents

Statistics: Introduction

- In the modern world of computers and information technology, the importance of statistics is very well recogonised by all the disciplines.
- Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Commerce, Biology, Medicine, Industry, planning, education and so on.

Definition of Statistics

- Statistics is defined differently by different authors over a period of time
- Statistics are numerical statement of facts in any department of enquiry placed in relation to each other. - A.L. Bowley
- Statistics may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis. It is clear that the definition of statistics by Croxton and Cowden is the most scientific and realistic one. According to this definition there are four stages:Collection of Data,Presentation of data,Analysis of data and Interpretation of data. - Croxton and Cowden:

Data Analysis

- The data can be collected in connection with time or geographical location or in connection with time and location.
- Any statistical data can be classified under two categories depending upon the sources utilized.
- Primiary data
- Secondary data.

Primiary data

Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by survey conducted by individuals or research institution or any organisation

Example

If a researcher is interested to know the impact of noonmeal scheme for the school children, he has to undertake a survey and collect data on the opinion of parents and children by asking relevant questions. Such a data collected for the purpose is called primary data.

Methods for Collecting Primary data

The primary data can be collected by the following five methods.

1. Direct personal interviews
2. Indirect Oral interviews
3. Information from correspondents
4. Mailed questionnaire method
5. Schedules sent through enumerators.

Secondary data

Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency.

Frequency distribution

Frequency distribution is a series when a number of observations with similar or closely related values are put in separate bunches or groups, each group being in order of magnitude in a series. It is simply a table in which the data are grouped into classes and the number of cases which fall in each class are recorded. It shows the frequency of occurrence of different values of a single Phenomenon.

Frequency distribution

A frequency distribution is constructed for three main reasons:

1. To facilitate the analysis of data.
2. To estimate frequencies of the unknown population distribution from the distribution of sample data and
3. To facilitate the computation of various statistical measures.

Raw data or Ungrouped data

The statistical data collected are generally raw data or ungrouped data.

Example

Let us consider the daily wages (in Rs) of 30 labourers in a factory. 800, 700, 550, 500, 600, 650, 400, 300, 800, 900, 750, 450, 350, 650, 700, 800, 820, 550, 650, 800, 600, 550, 380, 650, 750, 850, 900, 650, 450, 750.

Discrete (or) Ungrouped frequency distribution

In this form of distribution, the frequency refers to discrete value. Here the data are presented in a way that exact measurement of units are clearly indicated.

Example

In a survey of 40 families in a village, the number of children per family was recorded and the following data obtained.

1	0	3	2	1	5	6	2
2	1	0	3	4	2	1	6
3	2	1	5	3	3	2	4
2	2	3	0	2	1	4	5
3	3	4	4	1	2	4	5

Discrete frequency distribution.

Represent the data in the form of a discrete frequency distribution.

Number of Childern	Frequency
0	3
1	7
2	10
3	8
4	6
5	4
6	2
Total	40

Continuous frequency distribution

In this form of distribution refers to groups of values. This becomes necessary in the case of some variables which can take any fractional value and in which case an exact measurement is not possible. Hence a discrete variable can be presented in the form of a continuous frequency distribution.

Example

Wage distribution of 100 employees

Weekly wages (Rs)	Number of employees
50-100	4
100-150	12
150-200	22
200-250	33
250-300	16
300-350	8
Total	100

MEASURES OF CENTRAL TENDENCY

- In the study of a population with respect to one in which we are interested we may get a large number of observations.
- It is not possible to grasp any idea about the characteristic when we look at all the observations.
- So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic.
- Such representative number can be a central value for all these observations.
- This central value is called a measure of central tendency or an average or a measure of locations.

MEASURES OF CENTRAL TENDENCY

There are five averages.

1. Mean
2. Median
3. Mode
4. Geometric mean and
5. Harmonic mean.

Arithmetic mean or Mean

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. If the variable x assumes n values $x_1, x_2, \ldots, x_n$ then the mean, $\bar{x}$ is given by

Arithmetic mean or Mean

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 +, \ldots, +x_n}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i\end{aligned}$$

Problem

A student's marks in 5 subjects are 75, 68, 80, 92, 56. Find his average mark.

Grouped Data:Discrete series

The mean for grouped data is obtained from the following formula:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i$$

where x_i=the given individual data value
f_i=the frequency of individual class
N=the sum of the frequencies or total frequencies.

Problem

Given the following frequency distribution, Calculate the arithmetic mean

Marks	64	63	62	61	60	59
Number of Students	8	18	12	9	7	6

Grouped Data:Continuous series

The mean for grouped data is obtained from the following formula:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i$$

where x_i=the mid-point value of individual class

f_i=the frequency of individual class

N=the sum of the frequencies or total frequencies.

Problem

Following is the distribution of persons according to different income groups. Calculate arithmetic mean.

Income Rs.(100)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of persons	6	8	10	12	7	4	3

Solution

Income C.I	Number of Students (f)	Mid x	fx
0-10	6	05	030
10-20	8	15	120
20-30	10	25	250
30-40	12	35	420
40-50	7	45	315
50-60	4	55	220
60-70	3	65	195
Total	**N=50**		**1550**

Solution

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i$$
$$= \frac{1550}{50}$$
$$= 31$$

Positional Averages

- These averages are based on the position of the given observation in a series, arranged in an ascending or descending order.
- Median
- Mode

Median

The median is that value of the variate which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median.

- **Ungrouped or Raw data :**
  Arrange the given values in the increasing or decreasing order. If the number of values are odd, median is the middle value. If the number of values are even, median is the mean of middle two values.
  Formula : $\left(\frac{n+1}{2}\right)^{th}$ value, where n is the number of observations.

Problem

Find median for the following data 25, 18, 27, 10, 8, 30, 42, 20, 53.

- **Solution**
  Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53.
  The middle value is the 5th value.
  That is 25 is the median.

Solution

Using Formula.
Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53.

- Here $n = 9$

$$
\begin{aligned}
Median &= \left(\frac{n+1}{2}\right)^{th} value \\
&= \left(\frac{9+1}{2}\right)^{th} value \\
&= \left(\frac{10}{2}\right)^{th} value \\
&= 5^{th} value \\
&= 25
\end{aligned}
$$

Problem

Find median for the following data 5, 8, 12, 30, 18, 10, 2, 22

- **Solution**

  Arranging the data in the increasing order 2, 5, 8, 10, 12, 18, 22, 30
  Here median is the mean of the middle two values (ie) mean of (10,12)
  That is $\left(\frac{10+12}{2}\right) = 11$ is the median.

Solution

Using Formula. Arranging the data in the increasing order 2, 5, 8, 10, 12, 18, 22, 30. Here $n = 8$.

$$
\begin{aligned}
Median &= \left(\frac{n+1}{2}\right)^{th} value \\
&= \left(\frac{8+1}{2}\right)^{th} value \\
&= \left(\frac{9}{2}\right)^{th} value \\
&= 4.5^{th} value \\
&= 4^{th} value + \frac{1}{2}\left[5^{th} value - 4^{th} value\right] \\
&= 10 + \frac{1}{2}\left[12 - 10\right] \\
&= 10 + \frac{1}{2}\left[2\right] = 11
\end{aligned}
$$

Problem

The following table represents the marks obtained by a batch of 10 students in certain class tests in statistics and English.

Students ID	Marks Statistics	Marks English
1	53	57
2	55	45
3	52	24
4	32	31
5	30	25
6	60	84
7	47	43
8	46	80
9	35	32
10	28	72

Indicate in which subject is the level of knowledge higher?

Solution

The marks in the two subjects are first arranged in increasing order as follows:
Marks in Statistics: 28, 30, 32, 35, 46, 47, 52, 53, 55, 60
Marks in English: 24, 25, 31, 32, 43, 45, 57, 72, 80, 84.
Here $n = 10$.

- Median(Statistics) is the mean of the middle two values (ie) mean of (46,47)
- That is $\left(\frac{46+47}{2}\right) = 46.5$ is the median(Statistics).
- Median(English) is the mean of the middle two values (ie) mean of (43,45)
  That is $\left(\frac{43+45}{2}\right) = 44$ is the median(English).
- Therefore the level of knowledge in Statistics is higher than that in English.

Median-Grouped Data

- Cumulative frequency (cf)
  Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the pervious classes, ie adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

Median-Grouped Data:Discrete Series

Procedure for calculating median is given as follows:

- Find cumulative frequencies
- Find $\left(\frac{N+1}{2}\right)$
- See in the cumulative frequencies the value just greater than $\left(\frac{N+1}{2}\right)$
- Then the corresponding value of x is median.

Problem

Given the following frequency distribution, Calculate the median.

Marks	*64*	*63*	*62*	*61*	*60*	*59*
Number of Students	*8*	*18*	*12*	*9*	*7*	*6*

Solution

Here $N = 60$ *and* $\left(\frac{N+1}{2}\right) = 30.5$.

Marks x	Number of Students (f)	cf
64	8	8
63	18	26
62	12	38
61	9	47
60	7	54
59	6	60

Therefore, in the cumulative frequencies the value just greater than $\left(\frac{N+1}{2}\right) = 30.5$ *is 38. Then the corresponding value of* x *is median=62.*

Median-Grouped Data:Continous series

Procedure for calculating median is given as follows:

- Find cumulative frequencies
- Find $\frac{N}{2}$
- See in the cumulative frequencies the value just greater than $\frac{N}{2}$
- Then the corresponding class interval is called Median class.
- Then apply the formula:

$$Median = l + h\left(\frac{\frac{N}{2} - cf}{f}\right)$$

Where l= Lower limit of the median class
cf= cumulative frequency preceding the median class
h= width of the median class
f= frequency in the median class.
N=Total frequency.

Problem

Following is the distribution of persons according to different income groups. Calculate median.

Income in Rs.100	20-30	30-40	40-50	50-60	60-70
Number of persons	3	5	20	10	5

Solution

Here $N = 43$. Then $\frac{N}{2} = 21.5$

Income c.I	Number of persons (f)	cf
20-30	3	3
30-40	5	8
40-50	20	28
50-60	10	38
60-70	5	43

Therefore,in the cumulative frequencies the value just greater than $\frac{N}{2} = 21.5$ is 28. Then the corresponding class interval (ie) median class is $40 - 50$.

Solution

Then $Median = l + h\left(\frac{\frac{N}{2} - cf}{f}\right)$, where

Lower limit of the median class $l = 40$

Cumulative frequency preceding the median class $cf = 8$

Width of the median class $h = 10$

Frequency in the median class $f = 20$ and Total frequency $N = 43$.

$$
\begin{aligned}
Median &= l + h\left(\frac{\frac{N}{2} - cf}{f}\right) \\
&= 40 + 10\left(\frac{\frac{43}{2} - 8}{20}\right) \\
&= 40 + \frac{13.5}{2} \\
&= 40 + 6.75 = 46.75
\end{aligned}
$$

Median of wage is Rs.4675.

Problem

The following data attained from a garden records of certain period Calculate the median weight of the apple.

Weight in grams	*Number of Apples*
410-420	*14*
420-430	*20*
430-440	*42*
440-450	*54*
450-460	*45*
460-470	*18*
470-480	*7*

Mode is defined as the value which occurs most frequently in a data set. The mode obtained may be two or more in frequency distribution.

- Ungrouped or Raw Data

  The mode is defined as the value which occurs frequently in a data set.

Problem

A teacher was conducted a test for test 20 students and their marks are recorded 90, 70, 50, 30, 40, 86, 65, 73, 68, 90, 90, 10, 73, 25, 35, 88, 67, 80, 74, 46. Find the mode value of marks.

Solution

Mode is 90.

Problem

A doctor who checked 9 patient's sugar level is recorded 80, 112, 110, 115, 124, 130, 100, 90, 150, 180. Find the mode value of the sugar level.

Solution

Since every value appear only once in the data set, it follows that there is no mode.

Problem

Compute mode value for the following observations 2, 7, 10, 12, 10, 19, 2, 11, 3, 12.

Solution

Here, the values 2, 10 and 12 are appear twice in the data set. Therefore, the modes are 2, 10 and 12.

It is clear that mode may not exist or mode may not be unique.

Mode

- Discrete series

  For discrete frequency distribution, mode is the value of the variable corresponding to the maximum frequency.

Problem

Calculate the mode from the following data

Days of Confinement	*6*	*7*	*8*	*9*	*10*
Number of patients	*4*	*6*	*7*	*5*	*2*

Solution

Here, 7 is the maximum frequency, hence the value of x corresponding to 7 is 8. Therefore,the mode is 8.

- Continuous series

The mode or modal value of the distribution is that value of the variate for which the frequency is maximum. It is the value around which the items or observations tend to be most heavily concentrated. The mode is computed by the formula.

$Mode = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c$, where

l is the lower limit of the modal class

f_1 is the frequency of the modal class

f_0 is the frequency of the class preceding the modal class

f_2 is the frequency of the class succeeding the modal class

c is the width of the class limit.

Remark

- *If* $2f_1 - f_0 - f_2$ *comes out to be zero, then mode is obtained by the following formula taking absolute differences.*
  $Mode = l + \left(\frac{f_1 - f_0}{|f_1 - f_0| + |f_1 - f_2|}\right) \times c$
- *If mode lies in the first class interval, then* f_0 *is taken as zero.*
- *The computation of mode poses problem when the modal value lies in the open-ended class.*

Determination of Modal class

For a frequency distribution modal class corresponds to the class with maximum frequency. But in any one of the following cases that is not easily possible.

- If the maximum frequency is repeated.
- If the maximum frequency occurs in the beginning or at the end of the distribution
- If there are irregularities in the distribution, the modal class is determined by the method of grouping.

Method of Grouping

We prepare a grouping table with 6 columns.

1. In column I, we write down the given frequencies.
2. Column II is obtained by combining the frequencies two by two.
3. Leave the first frequency and combine the remaining frequencies two by two and write in column III.
4. Column IV is obtained by combining the frequencies three by three.
5. Leave the first frequency and combine the remaining frequencies three by three and write in column V.
6. Leave the first and second frequencies and combine the remaining frequencies three by three and write in column VI.

Mark the highest frequency in each column. Then form an analysis table to find the modal class. Use the mode formula to calculate the modal value.

Problem

The following data relates to the daily income of families in an urban area. Find the modal income of the families.

Income in Rs.	Number of persons
0-100	5
100-200	7
200-300	12
300-400	18
400-500	16
500-600	10
600-700	5

Solution

Here the maximum frequency is 18, therefore the modal clss is 300-400.

Income in Rs.	Number of persons
0-100	5
100-200	7
200-300	12
300-400	18
400-500	16
500-600	10
600-700	5

The lower limit of the modal class $l = 300$
The frequency of the modal class $f_1 = 18$
The frequency of the class preceding the modal class $f_0 = 12$
The frequency of the class succeeding the modal class$f_2 = 16$
The width of the class limit $c = 100$.

Solution Continued...

$$
\begin{aligned}
Mode &= l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c \\
&= 300 + \left(\frac{18 - 12}{2(18) - 12 - 16}\right) \times 100 \\
&= 300 + \frac{6}{8} \times 100 \\
&= 300 + 0.75(100) \\
&= 375.
\end{aligned}
$$

Problem

Calculate mode for the following frequency distribution:

Income in Rs.	Number of persons
0-5	9
5-10	12
10-15	15
15-20	16
20-25	17
25-30	15
30-35	10
35-40	13

Solution

Since there are irregularities in the distribution, because the frequeny values are incresing order 9,12,15,16,17 then gradually decreasing 15,10 but the next value 13 is increasing from the previous value 10. Therefore, use grouping method to determine the Modal class.

Solution Continued

Class	I	II	III	IV	V	VI
0-5	9	21		36		
5-10	12		27		43	
10-15	15	31				48
15-20	16		33	48		
20-25	17	32			42	
25-30	15		25			38
30-35	10	23				
35-40	13					

Solution Continued

Class	I	II	III	IV	V	VI	Total
0-5							
5-10					1		1
10-15					1	1	2
15-20			1	1	1	1	4
20-25	1	1	1	1		1	5
25-30		1		1			2
30-35							
35-40							

The maximum number of 1's appear in the class interval is 20-25.
Therefore the Modal class is 20-25.

Solution Continued

The lower limit of the modal class $l = 20$
The frequency of the modal class $f_1 = 17$
The frequency of the class preceding the modal class $f_0 = 16$
The frequency of the class succeeding the modal class $f_2 = 15$
The width of the class limit $c = 5$.

$$\begin{aligned}
Mode &= l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c \\
&= 20 + \left(\frac{17 - 16}{2(17) - 16 - 15}\right) \times 5 \\
&= 20 + \frac{1}{3} \times 5 \\
&= 20 + 0.33(5) \\
&= 21.65
\end{aligned}$$

Remark

Karl Pearson has expressed this relationship as follows
Mode=3 Median - 2 Arithmetic Mean

Measures of Dispersion

- The various measures of averages give a single number as the representative of the whole data.
- But they do not give how the observations are scattered about the average.
- Example

  Two distributions giving the weekly wages if 200 persons may have the same mean value,say Rs. 100.

  In one distributionn, most of the observatons may be centered around the mean value 100, a few others away from 100

  In another distribution a large number of observations may be above 150 and another set of large number of observations may be below 50 and only a few between 50 and 100 and still mean may be 100.

From this example,

- These two distributions with the same mean are not identical.
- In one, the items are nearer to the mean and in the other they are spread away from the mean.
- Two distributions may also have same median. But the deviations of the observations from the median may be different type in the two distributions.

  To study this aspect of distributions,another characteristic called dispersion.

- Example

Student-1 Marks	Student-2 Marks
68	83
75	85
65	82
67	20
60	65
Total=335	Total=335

Here mean=67 and hence student-1 has less variation than student-2.
Here Chartertistic: less variation.

There are two kind of measures of dispersion

1. Absolute measure of dispersion
2. Relative measure of dispersion

Absolute measure of dispersion

Absolute measure of dispersion indicates the amount of variations in a set of values in terms of units of observation.

Example

When rainfalls on differ days are available in mm, any absolute measure of dispersion gives the variation rainfall in mm.

Relative measure of dispersion

Relative measures of dispersion are free from the units of measurements of the observations.

They are pure numbers. They are used to compare the variation in two or more sets, which are having differnent units of measuremet of observation.

Absolute measure

1. Range
2. Quartile deviation
3. Mean deviation
4. Standard deviation

Relative measure

1. Co-efficient of Range
2. Co-efficient of Quartile deviation
3. Co-efficient of Mean deviation
4. Co-efficient of variation

- Range and Co-efficient of Range
  Range
  Range=L-S, where L-Largest value, S-Smallest value.
  Co-efficient of Range
  Co-efficient of Range=$\frac{L-S}{L+S}$, where L-Largest value, S-Smallest value.
- Find L and S for Continuous Series
  Method-I
  L-Upper boundary of the highest class
  S-Lower boundary of the lowest class
  Method-II
  L-Mid value of the highest class
  S-Mid value of the lowest class

Quartiles

The quartiles divided the distributions in four equal parts.

Quartiles

- The first quartile(Q1) marks off the first one-fouth
- The second quartile divides the distribution into two halves (median)(Q2).
- The third quartile(Q3) marks off the three-fourth.

Quartiles:Ungrouped data

- To arrange the given data in the increasing order
- Use the formula for $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ value
- Use the formula for $Q_2 = \left(\frac{n+1}{2}\right)^{th}$ value
- Use the formula for $Q_3 = \left(\frac{3(n+1)}{4}\right)^{th}$ value

Problem

Compute quartiles for the data given below 25,18,30,8,15,5,10,35,40,45.

Solution

Arranging the data in increasing order: 5,8,10,15,18,25,30,35,40,45

$Q_1 = \left(\frac{n+1}{4}\right)^{th}$ *value*

$= (2.75)^{th}$ *value*

$= 2^{nd} value + \frac{3}{4}(3^{rd} value - 2^{nd} value)$

$= 8 + \frac{3}{4}(10 - 8)$

$= 8 + \frac{3}{2}$

$= 9.5$

Solution Cont...

Arranging the data in increasing order: 5,8,10,15,18,25,30,35,40,45

$Q_3 = \left(\frac{3(n+1)}{4}\right)^{th}$ value

$= (8.25)^{th}$ value

$= 8^{th} value + \frac{1}{4}(9^{th} value - 8^{th} value)$

$= 35 + \frac{1}{4}(40 - 35)$

$= 35 + \frac{5}{4}$

$= 36.25$

Quartiles:Grouped data-Discrete series

- Find Cumulative frequencies(c.f)
- Find $\left(\frac{N+1}{4}\right)^{th}$ value, where N is the total frequency
- See in the cumulative frequencies, the value just greater than $\frac{N+1}{4}$, then the corresponding value of x is Q_1.
- Find $\left(\frac{3(N+1)}{4}\right)^{th}$ value,
- See in the cumulative frequencies, the value just greater than $\frac{3(N+1)}{4}$, then the corresponding value of x is Q_3.

Problem

Compute quartiles for the data given below

x	f
5	4
8	3
12	2
15	4
19	5
24	2
30	4

Quartiles:Grouped data-Continuous series

- Find Cumulative frequencies(c.f)
- Find $\frac{N}{4}$ value, where N is the total frequency
- See in the cumulative frequencies, the value just greater than $\frac{N}{4}$, then the corresponding class interval is called first quartile class.
- Find $\frac{3N}{4}$ and see in the cumulative frequencies, the value just greater than $\frac{3N}{4}$, then the corresponding class interval is called third quartile class.
- Then apply the respective formula

$$Q_1 = l_1 + h_1 \left(\frac{\frac{N}{4} - cf_1}{f_1} \right)$$

$$Q_3 = l_1 + h_3 \left(\frac{\frac{3N}{4} - cf_3}{f_3} \right)$$

Where $l_1(l_3)$= Lower limit of the first(third) quartile class

$cf_1(cf_3)$= cumulative frequency preceding the first(third) quartile class

$h_1(h_3)$= width of the first(third) quartile class

$f_1(f_3)$= frequency of the first(third) quartile class

N=Total frequency.

Problem

The following series relates to the marks secured by students in an examination. Compute quartiles.

Marks	Number of students
0-10	11
10-20	18
20-30	25
30-40	28
40-50	30
50-60	33
60-70	22
70-80	15
80-90	12
90-100	10

Deciles

- These are the values, which divide the total nmber of observation into 10 equal parts. Therefore, there are 9 deciles, $D_1, D_2, \ldots, D_9$.
- Ugrouped Data
  To arrange the given data in the increasing order
  $D_i = \left(\frac{i(n+1)}{10}\right)^{th}$ value, where $i = 1, 2, 3, \ldots, 9$.

Problem

Compute D_5 and D_8 from the data given below 5,24,36,12,20,8

Solution

Arranging the given data in the increasing order 5,8,12,20,24,36

$D_5 = \left(\frac{5(n+1)}{10}\right)^{th}$ *value*

$= (3.5)^{th}$ *value*

$= 3^{rd} value + \frac{1}{2}(4^{th} value - 3^{rd} value)$

$= 12 + \frac{1}{2}(20 - 12)$

$= 12 + 4$

$= 16$

Deciles for grouped data-Discrete series

- Find Cumulative frequencies(c.f)
- Find $\frac{iN}{10}$, where N is the total frequency
- See in the cumulative frequencies, the value just greater than $\frac{iN}{10}$, then the corresponding value of x is D_i.

Deciles for Grouped data-Continuous series

- Find Cumulative frequencies(c.f)
- Find $\frac{iN}{10}$ value, where N is the total frequency
- See in the cumulative frequencies, the value just greater than $\frac{iN}{10}$, then the corresponding class interval is called deciles class.
- Then apply the respective formula

$$D_i = l_i + h_i \left(\frac{\frac{iN}{10} - cf_i}{f_i} \right), \text{ where } i = 1, 2, 3, 4, 5, 6, 7, 8, 9.$$

Where l_i= Lower limit of the deciles class
cf_i= cumulative frequency preceding the deciles class
h_i= width of the deciles class
f_i= frequency of the fdeciles class
N=Total frequency.

Problem

The following series relates to the marks secured by students in an examination. Calculate D_3 and D_7.

Marks	Number of students
0-10	5
10-20	7
20-30	12
30-40	16
40-50	10
50-60	8
60-70	4

Percentiles

The percentile values divide the distribution into 100 parts each containing 1 percent of the cases. It is denoted by P_k.
The percentile P_k is that value of the variable up to which lie exactly k% of the total number of observations.
Relationship $P_{25} = Q_1, P_{50} = D_5 = Q_2 = median$ and $P_{75} = Q_3$.

Percentiles

- Ugrouped Data
  To arrange the given data in the increasing order
  $P_i = \left(\frac{i(n+1)}{100}\right)^{th}$ value, where $i = 1, 2, 3, \ldots, 99$.
- Problem Compute P_{15} and P_{42} from the data given below
  5,24,36,12,20,8.

Percentiles for grouped data-Discrete series

- Find Cumulative frequencies(c.f)
- Find $\frac{iN}{100}$, where N is the total frequency
- See in the cumulative frequencies, the value just greater than $\frac{iN}{100}$, then the corresponding value of x is P_i.

Percentiles for Grouped data-Continuous series

- Find Cumulative frequencies(c.f)
- Find $\frac{iN}{100}$ value, where N is the total frequency
- See in the cumulative frequencies, the value just greater than $\frac{iN}{100}$, then the corresponding class interval is called percentiles class.
- Then apply the respective formula

$$P_i = l_i + h_i \left(\frac{\frac{iN}{100} - cf_i}{f_i} \right), \text{ where } i = 1, 2, 3, \ldots, 99.$$

Where l_i= Lower limit of the decciles class

cf_i= cumulative frequency preceding the decciles class

h_i= width of the decciles class

f_i= frequency of the decciles class

N=Total frequency.

Problem

The following series relates to the marks secured by students in an examination. Calculate P_{53}.

Marks	Number of students
0-5	5
5-10	8
10-15	12
15-20	16
20-25	20
25-30	10
30-35	4
35-40	4
40-45	3

Quartile deviation and Co-efficient of Quartile

- Quartile deviation(Q.D)

  Quartile deviation is half of the difference between the first and third quartiles.

  $$Q.D = \frac{Q_3 - Q_1}{2}.$$

- Co-efficient of Quartile

  Co-efficient of $Q.D = \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$.

Thank you