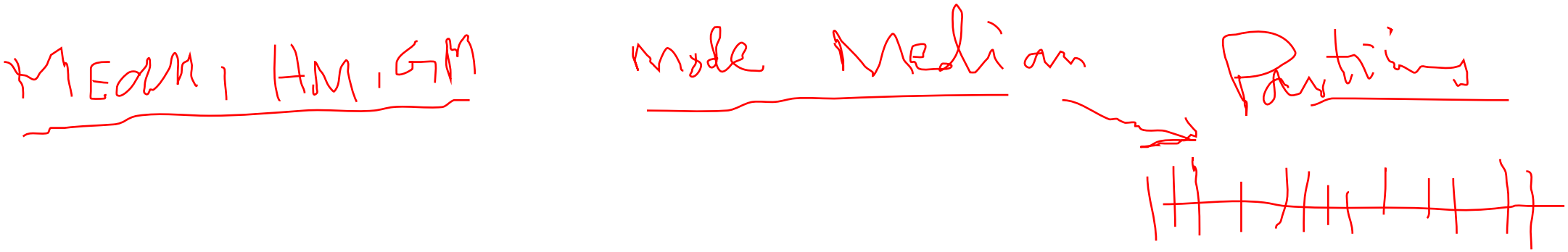
Lecture No. 5

Introduction To Statistics, Statistics And Probability

Dr. Shabbir Ahmad

Assistant Professor,
Department of Mathematics,
COMSATS University
Islamabad, Wah Campus



Measures Of Non Central Tendency

Mode, Media, Quartiles, Deciles and Percentiles

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:12 PM

In this lecture

- Mode of Ungrouped and Grouped data
- Merits and Demerits of Mode
- Median of Ungrouped and Grouped data
- Merits and Demerits of Median
- Quartile, Decile and Percentiles of Ungrouped and Grouped data

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:12 PM

Mode

The most frequent or most repeated value in in the data is called mode. Mode can be calculated by just by inspection. For example the mode of the values 2, 4, 3, 4, 3, 7, 6, 3 is 3, as 3 is repeated most of the times

It is classified as unimodal, bimodal, trimodal or mulitimodal

Unimodal is a distribution of scores that consists of only one mode.

Bimodal is a distribution of scores that consists of two modes.

Trimodal is a distribution of scores that consists of three modes

Multimodal is a distribution of scores that consists of more than two modes.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Mode For Grouped Data

For grouped data the mode is obtained by using formula:

$$Mode = l + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

Where,

l **= lower class boundary of modal class,**

f_m **= maximum frequency of the distribution,**

f_1 **= frequency prior to the maximum frequency,**

f_2 **=frequency lower to the maximum frequency,**

h **= class interval of the modal class,**

and

the modal class is the class containing maximum frequency of the distribution.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Question: Calculate the mode of the following frequency distribution.

Classes	110 – 119	120 – 129	130 – 139	140 – 149	150 – 159	160 – 169	170 – 179	180 – 189	190 – 199	200 – 209	210 – 219
f	1	4	17	28	25	18	13	6	5	2	1

$$Mode = l + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

Where, $l = 139.5, f_m = 28, f_1 = 17, f_2 = 25$ and $h = 10$

$$Mode = 139.5 + \frac{(28-17)}{(28-17)+(28-25)} \times 10$$

$$Mode = 139.5 + \frac{11}{11+3} \times 10$$

$$Mode = 139.5 + 0.785714(10) = 139.5 + 7.85714$$

$$Mode = 147.35714$$

Classes	f	C.B
110 – 119	1	109.5 – 119.5
120 – 129	4	119.5 – 129.5
130 – 139	$17 = f_1$	129.5 – 139.5
140 – 149	$28 = f_m$	$l = 139.5$ – 149.5
150 – 159	$25 = f_2$	149.5 – 159.5
160 – 169	18	159.5 – 169.5
170 – 179	13	169.5 – 179.5
180 – 189	6	179.5 – 189.5
190 – 199	5	189.5 – 199.5
200 – 209	2	199.5 – 209.5
210 – 219	1	209.5 – 219.5
Total	120	

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Merits and Demerits af Mode

Merits:

- **It is easy to calculate and simple to understand**
- **It is not affected by extreme values**
- **It can be useful for qualitative data**

Demerits:

- **It is not well defined by mathematical formulae**
- **It is not based on all observation**
- **It is stable for large number of values and my be ill-defined for small number of values**
- **Some time a distribution may have no mode at all and some times more than one mode.**

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Median

The MEDIAN, is the middle value of the sample when the data are arranged in order according to size.

Connor has defined as " The median is that value of the variable which divides the group into two equal parts, one part comprising of all values greater, and the other, all values less than median"

It is also known as the middle score or the 50^{th} percentile.

Fifty percent (50%) of the ordered data values lies below the median value and 50% lies above the median value.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Median of UG Data

Median is the most middle value in the array data, i.e., when the data are arranged in ascending or descending order, median is the middle value if there is odd number of values and is the mean of two middle values when the data is even number of values.

For example,

The median of the data 2,3,5,9,11 (odd number of values) is 5.

The median of the data 2,3,5,9,11, 13 (even number of values) is $\frac{(5+9)}{2} = 7$

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Median for Grouped Data

$$M = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

Where,

***l*= Lower class boundary of median class,**

***h*= Class interval of median class,**

***f*= frequency of median class,**

***n*= total no. of observation in a data set or sum of all the frequencies,**

***C*= Cumulative frequency of the class prior to the median class,**

and the median class is the class containing $(\frac{n}{2}th)$ observations (identify from cumulative frequency)

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Question: Calculate the median of the following frequency distribution

$Classes$	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180-189	190-199	200-209	210-219
f	1	4	17	28	25	18	13	6	5	2	1

Classes	f	C. B	Cmf
110 – 119	1	109.5 – 119.5	1
120 – 129	4	119.5 – 129.5	5
130 – 139	17	129.5 – 139.5	22
140 – 149	28	139.5 – 149.5	$50 = C$
150 – 159	**$25 = f$**	**$l = 149.5 – 159.5$**	**75**
160 – 169	18	159.5 – 169.5	93
170 – 179	13	169.5 – 179.5	106
180 – 189	6	179.5 – 189.5	112
190 – 199	5	189.5 – 199.5	117
200 – 209	2	199.5 – 209.5	119
210 – 219	1	209.5 – 219.5	120
	$120 = n = \sum f$		

$\frac{n}{2} = \frac{\sum f}{2} = 60$

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Solution: **The first step is to identify the median class, here**

$$Median\ Class = Class\ containing\left(\frac{n}{2} = \frac{120}{2} = 60\right) th\ observation.$$

The mentioned class (150-159) is the median class as $cmf = 75$ shows that 75 observations are studied under this class i.e. this class contains 60^{th} observation of the data.

Now using formula:

$$M = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

Where, $l = 149.5, h = 10, f = 25, C = 50\ and\ n = 120$

$$M = 149.5 + \frac{10}{25}\left(\frac{120}{2} - 50\right) = 149.5 + 0.4(60 - 50) = 149.5 + 0.4(10)$$

$$M = 149.5 + 4 = 153.5$$

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Merits and Demerits of Median

Merits:

It is easy to calculate and simple to understand.

It is not affected by extreme values.

It is an appropriate average be used for extremely skewed distribution.

It can be located graphically.

Demerits:

It is not well defined by mathematical formulae.

It is not based on all observation.

It is stable for large number of values and my be ill-defined for small number of values.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

QUESTIONS FOR PRACTICE:

Question 1: The following sample data of the number of communications are taken from logs of communication with Distance Education students: 5, 9, 5, 23, 27, 55, 34, 7, 30, 15, 22, 60, 14, 52, 297, 8, 51, 15, 51, 35, 15, 39, 137, 43, 38, 14, 93, 7. Calculate the median and the mode.

Question 2: The following is the age distribution of 1000 persons working in an organization:

Age Group	Number of Persons
20-25	30
25-30	160
30-35	210
35-40	180
40-45	145
45-50	105
50-55	70
55-60	60
60-65	40

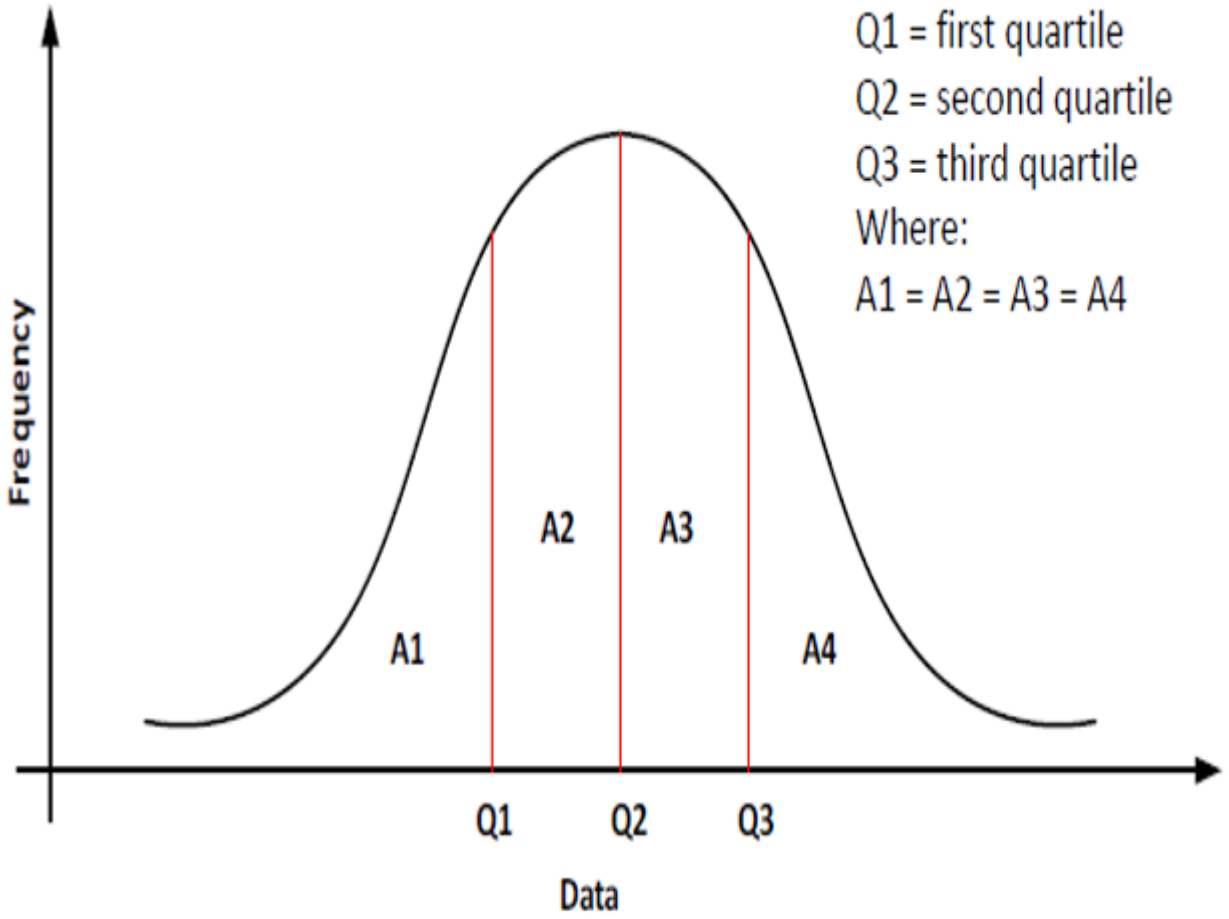
Calculate the median and the mode.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Quartiles

- Quartiles in statistics are values that divide your data into quarters, Quartiles divide the area under the curve into four equally parts.
- However, quartiles aren't shaped like pizza slices; Instead they divide your data into four segments according to where the numbers fall on the number line.
- The quartiles that divide a data set into four quarters are:

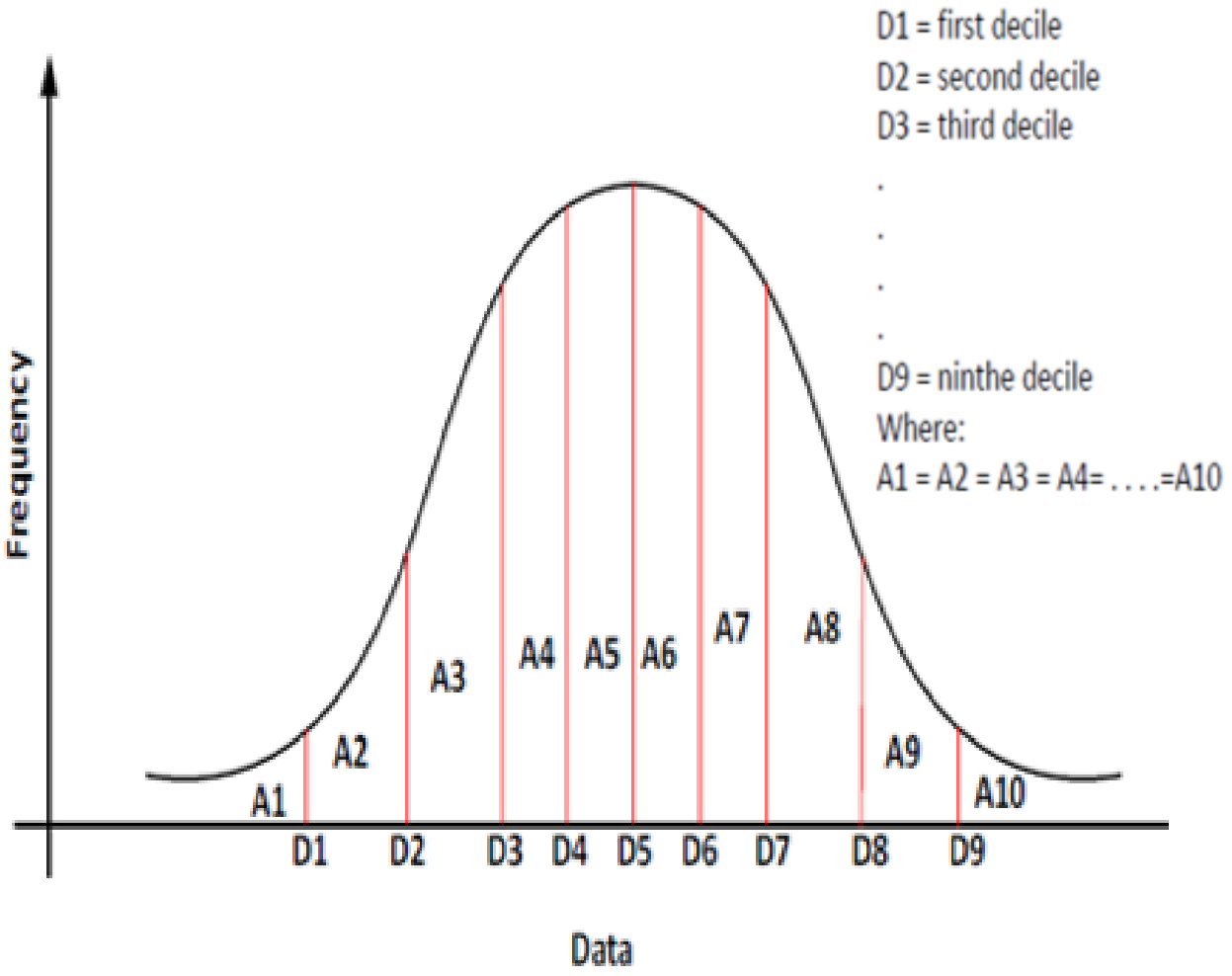
$$Q_1 = First\ Quartile$$
$$Q_2 = 2nd\ Quartile$$
$$Q_3 = 3rd\ Quartile$$



Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Deciles

- Decile is any of the nine values that divide the sorted data into ten equal parts, so that each part represents $1/10$ of the sample or population. Decile divide the area under the curve into ten equally pieces .
- The deciles that divide a data set into ten equal pieces are: $D_1, D_2, D_3, \ldots D_9$ respectively.



Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:05:13 PM

Percentile

A percentile (or a centile) is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations falls. Percentiles divide the area under the curve into 100 equally pieces .

The same procedure for division (as in quartile and decile) is done for finding percentiles for any frequency distributed curve. The percentiles that divide a data set into 100 equal pieces are:

P_1 = first percentile

P_2 = second percentile

P_3 = third percentile

.

P_{99} = ninety ninth percentile

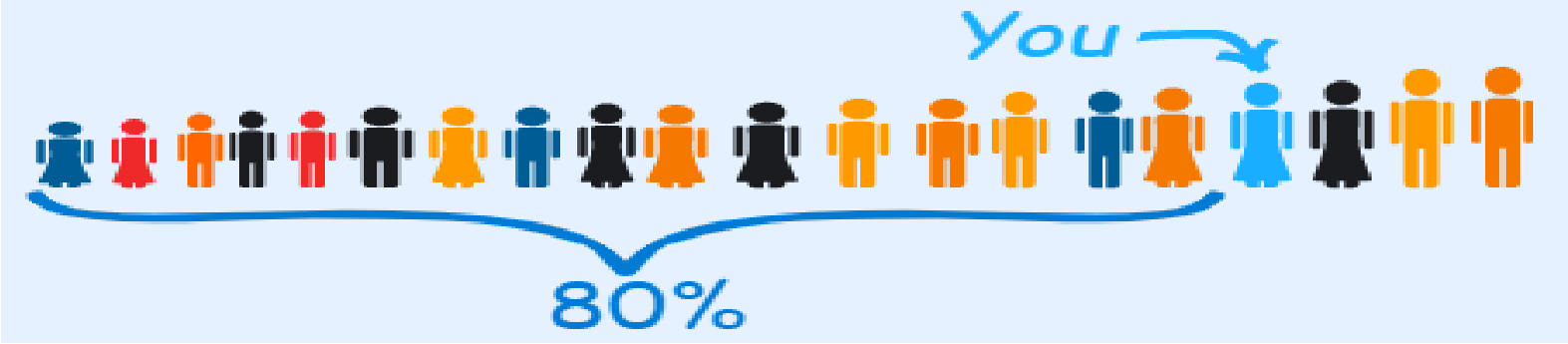
Where: A1 = A2 = A3 = A4= . . . .=A100

Example: You are the fourth tallest person in a group of 20

80% of people are shorter than you.

That means you are at the **80th percentile**.

If your height is 1.85m then "1.85m" is the 80th percentile height in that group.



Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

Formula & Examples

- To find the quartiles, or deciles, or percentiles we follow the same procedure to find the median.
- Arrangement the data in ascending form only.

Calculating Median (M)

M class = class with $(\frac{in}{2}th)$ value of the observation

$$M = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

Calculating Quartiles (Q_i)

Q_i class = class with $(\frac{in}{4}th)$ value of the observation

$$Q_i = l + \frac{h}{f}\left(\frac{in}{4} - C\right), \quad i = 1, 2, 3$$

Calculating Deciles (D_i)

D_i class = class with $(\frac{in}{10}th)$ value of the observation

$$D_i = l + \frac{h}{f}\left(\frac{in}{10} - C\right), \quad i = 1, 2, \ldots, 9$$

Calculating Percentiles (P_i)

P_i class = class with $(\frac{in}{100}th)$ value of the observation

$$P_i = l + \frac{h}{f}\left(\frac{in}{100} - C\right), \quad i = 1, 2, \ldots, 99$$

Where l, h, f, C are obtained by following the same procedure as followed in median for respective quartile, decile and percentile.

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

Example: Calculate Quartile-3, Decile-8, Percentile-20 from the following grouped data.

Classes	Frequency
2-4	3
4-6	4
6-8	2
8-10	1

Solution: 3rd Quartile:

Class	Frequency f	Cmf
2 - 4	3	3
4 - 6	4	7
6 - 8	2	9
8 - 10	1	10
---	---	---
	$n = 10$	

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

As Q_i class = class with $(\frac{in}{4}\,th)$ value of the observation, where $i = 3$

Q_3 class = class with $(\frac{3(10)}{4}\,th)$ value in Cmf column= class with (7.5th) value of the observation in Cmf

Column and it lies in the mentioned class $6 - 8$.

Using formula:

$$\boldsymbol{Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right)}$$

Where, $l = 6$ lower class boundary of Q_3 class , $h = 2$ class interval, $f = 2$ frequency of Q_3 class , $C = 7$ Cmf of prior class from Q_3 class

$$Q_3 = 6 + \frac{2}{2}\left(\frac{3(10)}{4} - 7\right) = 6 + 1(7.5 - 7) = 6 + 0.5 = 6.5$$

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

8TH Decile

Solution:

D_i class = class with $(\frac{in}{10}th)$ value of the observation, where $i = 7$

D_7 class = class with $(\frac{8(10)}{10}th)$ value of the observation in Cmf column= class with 8th observation in Cmf column and it lies in the mentioned column $6-8$, thus D_8 class is $6-8$.

Now using formula:

$$\boldsymbol{D_8 = l + \frac{h}{f}\left(\frac{8n}{10} - C\right)}$$

Class	Frequency f	Cmf
2 - 4	3	3
4 - 6	4	7
6 - 8	**2**	**9**
8 - 10	1	10
---	---	---
	$n = 10$	

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

Where, $l = 6$ lower class boundary of D_8 class , $h = 2$ class interval, $f = 2$ frequency of D_8 class , $C = 7$ Cmf of prior class from D_8 class

$$D_8 = 6 + \frac{2}{2}\left(\frac{8(10)}{10} - 7\right) = 6 + 1(8 - 7) = 6 + 1 = 7$$

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

20^{TH} Percentile

P_i class = class with $(\frac{in}{100}\,th)$ value of the observation, where i=20

P_{20} class = class with $(\frac{20(10)}{100}\,th)$ value of the observations= class with 2nd observation in Cmf column and it lies in the mentioned column $2-4$, thus P_{20} class is $2-4$.

Now using formula:

$$\boldsymbol{P_{20} = l + \frac{h}{f}\left(\frac{20n}{100} - C\right)}$$

$Class$	$Frequency$ f	Cmf
$2-4$	3	3
$4-6$	4	7
$6-8$	2	9
$8-10$	1	10
$---$	$---$	$---$
	$n = 10$	$--$

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

Where, $l = 2$ lower class boundary of P_{20} class , $h = 2$ class interval,

$f = 3$ frequency of P_{20} class , $C = 0$ Cmf of prior class from P_{20} class

$$P_{20} = 2 + \frac{2}{3}\left(\frac{20(10)}{100} - 0\right)$$

$$P_{20} = 2 + 0.667(2 - 0)$$

$$P_{20} = 2 + 1.333 = 3.3333$$

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

Questions for Practice

Question 1: The following sample data of the number of communications are taken from logs of communication with Distance Education students: 5, 9, 5, 23, 27, 55, 34, 7, 30, 15, 22, 60, 14, 52, 297, 8, 51, 15, 51, 35, 15, 39, 137, 43, 38, 14, 93, 7. Calculate $Q_1, Q_3, \mathrm{D}_5, \mathrm{D}_4, , \mathrm{P}_{50}, \mathrm{P}_{75}$.

Question 2: The following is the age distribution of 1000 persons working in an organization:

Calculate $Q_1, Q_3, \mathrm{D}_5, \mathrm{P}_{50}$.

Show that: $\mathrm{Median} = \mathrm{Q}_2 = \mathrm{D}_5 = \mathrm{P}_{50}$.

Age Group	Number of Persons
20-25	30
25-30	160
30-35	210
35-40	180
40-45	145
45-50	105
50-55	70
55-60	60
60-65	40

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM

ANY QUESTION

Dr. Shabbir Ahmad
Assistant Professor, Department of Mathematics, COMSATS University Islamabad, Wah Campus
Cell # 0323-5332733, 0332-5332733. Date: 25-Sep-21 12:54:27 PM