Unit 1 Basic Statistics

Data

Raw   Descrete   Continous

$x_1 + x_2 + x_3$

x	1	2	3	4
f	5	6	2	1

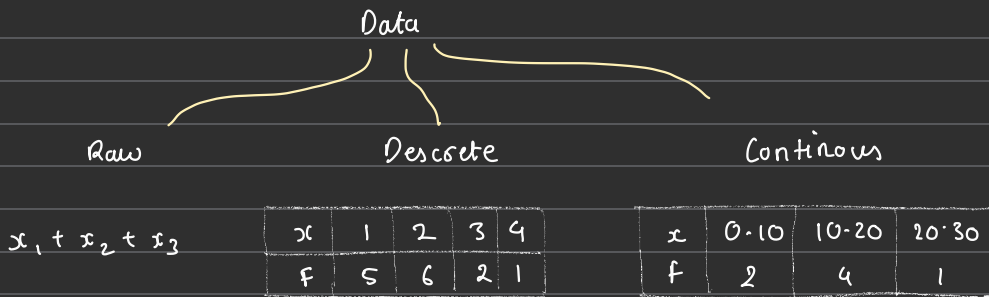| x	0-10	10-20	20-30
f	2	4	1

Statistics

     Statistics is the study of How to collect, organise, analyse and interpret the information about one or more population under investigation.

- ## Frequency

     The Number of Observations in a particular class interval is called a frequency of that class.

- ## Measure of Central Tendancy

     A single value which will be a representative of whole data and around which a large proportion of observation is clustered is called an average or measure of central tendancy.

Mean   Median   Mode

avg   middle   most
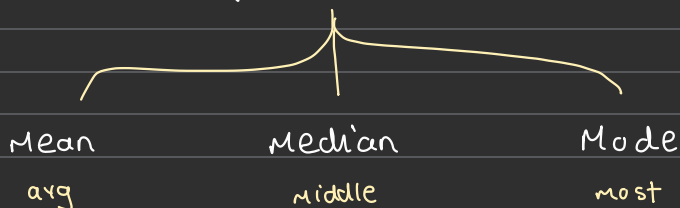
Mean

Represents the average of all the values for one variable in a dataset.

Median

Represents the Middle value when all the values are arranged in increasing order.

Mode

Most Common Value for a range.

* ## Mean

The Arithmatic Mean is the sum of observations divided by the number of Observations.

Case 1   Raw Data

If $x_1, x_2 \dots x_n$ are in observations, then by definitions of arithmetic mean

$$\overline{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n} \qquad \text{i.e} \qquad \frac{x_1 + x_2 \dots + x_n}{n}$$

The Notation for Arithmetic Mean is $\overline{x}$

Case II   Descrete frequency Distribution

Suppose $x_1, x_2 \ldots x_n$ are the values with $f_1, f_2 \ldots f_n$
frequencies Respectively, then

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 \ldots f_n x_n}{f_1 + f_2 \ldots f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Case III   Continous frequency Distribution

For Calculation, we make an assumption, that the frequency
is associated with the midpoint taking $x_1, x_2 \ldots x_n$
as the midpoint values.

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

where x_i
are midpoint
values

* lets Solve Some Examples

Q. The following Data Gives Monthly Salary of 10 employees
in an office. Calculate the Arithmatic Mean of the Salaries.

1780, 1760, 1690, 1750, 1840, 1920, 1100, 1810, 1050, 1950

→ We All Know, Arithmatic Mean = $\dfrac{\Sigma xi}{n}$

∴ $\dfrac{\text{Total Salary}}{10}$ = $\dfrac{15650}{10}$ ⇒ $\underline{\underline{1565}}$

Q. The Number of Telephone calls recieved in successive one minute is 293, intervals of an exchange are shown in the table.
Calculate Arithmatic Mean.

No. of Calls	frequency	fixi
0	10	0
1	28	28
2	35	70
3	45	135
4	65	260
5	52	260
6	32	192
7	12	84
8	14	112
	293	1141

→ Now, first we calculate fixi values ↗

now, using formula $\dfrac{\Sigma fixi}{\Sigma fi}$ ⇒ $\dfrac{1141}{293}$

⇒ 3.89   calls.

Q. Find the Arithmatic Mean for the following Data

Marks	no. of Students	Median (xi)	fi x'i
0 - 10	5	5	25
10 - 20	10	15	150
20 - 30	25	25	625
30 - 40	30	35	1050
40 - 50	20	45	900
50 - 60	10	55	550
	100		3300

Here, we first calculate median of every class x'_i

Then we calculate the $fi x_i$ values

Now, using formula $\dfrac{\Sigma \, fi \, x'_i}{\Sigma \, fi}$ =) $\dfrac{3300}{100} = 33$

Q. Given the Arithmatic Mean of the Data having 100 observations is 44.2
Find the Missing frequencies.

Class	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Freq.	5	12	a	20	b	10	4
Median	15	25	35	45	55	65	75
fixi	75	300	35a	900	55b	650	300

Calculating median x'_i & the $fixi$ values for given Data

Given   Mean $\sim$ 44.2
   Total observations = 100

Using formula : $\dfrac{\Sigma\, f_i x_i}{\Sigma\, f_i}$ = 44.2

$\Rightarrow$ $\dfrac{2225 + 35a + 55b}{100}$ = 44.2

$\Rightarrow$ $445 + 7a + 11b$ = 884

$$7a + 11b = 439 \quad\text{———}\quad ①$$

Now, from table we know total values = 100

$\therefore$ $a + b + 51 = 100$

$\Rightarrow$ $a + b = 49$ ——— ⑪

solving eq ① & eq ② we get.

$$\underline{\underline{a = 25}} \qquad \underline{\underline{b = 24}}$$

Q. The Arithmatic Mean is   5, find the value of x.

Variable	2	4	6	8
Freq.	$x-1$	$x+1$	$x+1$	$2x-5$
$f_i x_i$	$2x-2$	$4x+4$	$6x+6$	$16x-40$

$\rightarrow$ lets calculate $f_i x_i$ first

Given Mean = 5

Now, using formula $\quad \dfrac{\Sigma\, f_i x_i}{\Sigma\, f_i}$ $\Rightarrow$ $\dfrac{28x - 32}{5x - 4} = 5$

$\Rightarrow \quad 28x - 32 = 25x - 20$

$\Rightarrow \qquad 3x = 12$

$\qquad x = \underline{\underline{4}}$

* Median

Median of a Distribution is the value of the variable which Divides the Dataset into two Parts

It is the value Such that, the number of observations above it is equal to the no. of observations below it.

It is Also Called as Positional Average.

eg. 11, 12, 13, 14, 15, 16
$\qquad\qquad\qquad\quad \uparrow$
$\qquad\qquad\quad$ Median

○ Case I Raw Data

If the number of observations is odd, the Median is the Middle Value.

If the number of observations is even, the Median is the Arithmatic Mean of the two middle terms.

ODD → $\left(\dfrac{n+1}{2}\right)^{th}$ observation

EVEN → $\dfrac{\left(\dfrac{n}{2}\right)^{th} + \left(\dfrac{n}{2}+1\right)^{th}}{2}$ observation

Case II   Descrete frequency Distribution

Step 1   Compute the Cumulative frequency Column

Step 2   let $N = \Sigma f_i$, Calculate $N/2$

step 3   Consider the Cumulative frequency which is just greater than $N/2$, then the corresponding value of x is the median.

Case III   Continous frequency Distribution

Step 1   Compute the cumulative frequency Column.

Step 2   let $N = \Sigma f_i$, & calculate $N/2$

Step 3   consider the cumulative frequency which is just greater than $N/2$, then the corresponding class is called the Median Class.

finally, it is obtained using the formula.

$$Median = L + \frac{h}{f}\left[\frac{N}{2} - CF\right]$$

where   L = lower limit of Class

h = width i.e upper limit — Lower limit

f = frequency of the Median Class

CF = cumulating frequency of the class preceeding to median class.

* lets Solve Some Examples.

Q. following are the sales of vegetable Stores in 1000's of Rupees in a City.

$$13, \quad 15, \quad 18, \quad 19, \quad 20, \quad 24$$

→   Median → Raw Data = Even observations

$$\frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2} \quad \text{observation}$$

$$= \frac{18+19}{2} \quad \Rightarrow \quad 18.5$$

Q. Calculate the Median for the following Data

x	F	CF
1	3	3
3	6	9
5	8	17
7	12	29
9	16	45
11	16	61
13	15	76
15	10	86
17	5	91

→ 1. lets Calculate Cumulative frequency

2. Calculate $N/2$ => $\frac{\sum fi}{2}$ => $\frac{91}{2}$ = 45.5

3. Now,
the imediate higher value of CF > $N/2$
will give us the median Class.

∴ 61 > 45.5 ∴ the value of x at median days

Median => $\underline{\underline{11}}$

H.W

Q Calculate the Median for the following Data.

Income	5000	6000	6500	8000	9000	1000
No. of peep	16	24	26	30	20	5
CF	16	40	66	96	116	121

→ 1. Calculating CF of Data.

2. Calculating $N/2$ $\Rightarrow$ $\dfrac{\Sigma fi}{2}$ $\Rightarrow$ $\dfrac{121}{2}$ $\Rightarrow$ 60.5

3. The imediate highes CF $>$ $\dfrac{N}{2}$ $=$ 66 > 60.5

∴ we get the median Claus thus Median = 6500

Q. Find Median of Given Data

wages	no. of workeos	CF
10 - 20	5	5
20 - 30	8	13
30 . 40	13	26
40 - 50	10	36
50 - 60	8	44

→ 1. lets calculate the CF

2. Calculating $N/2$ = $\Sigma fi/2$ = $44/2$ = 22

3. finding Median claus = CF > 22

$\Rightarrow$ 26 > 22

Median Class = 30 . 40

using formula = $L + \dfrac{h}{f}\left[\dfrac{N}{2} - CF\right]$

$\Rightarrow$ $30 + \dfrac{10}{13}\left[22 - 13\right]$

$\Rightarrow$ 36.923 = Median

Q. Calculate the Median of the following Data.

fine	20-40	40-60	60-80	80-100	100- above
no. of Diff	7	22	39	39	13
CF	7	29	68	107	120

→ 1. Calculating Cf

2. $N/2 = \Sigma fi/2$ =) $120/2 = 60$

3. $CF > 60 = $ Median Class $= 60-80$

Using formula $= L + \dfrac{h}{f}\left[\dfrac{N}{2} - CF\right]$

=) $60 + \dfrac{10}{39}\left[120 - 29\right]$

Median =) 414.9

Q. Find the Value of a & b where the $\Sigma fi = 230$
Median = 46

Masks	10-20	20-30	30-40	40-50	50-60	60-70
students	12	30	a	65	b	43
CF	12	42	a+42	a+107	a+b+107	a+b+150

→ 1. Calculate CF

2. $N/2 = \Sigma fi/2 = 230/2$ =) 115

Now, $a+b+ 150 = 230$

$$a + b = 80 \qquad - \text{①}$$

Using formula $\quad L + \dfrac{h}{f} \left[\dfrac{N}{2} - CF \right]$

As $M = 46$, median Class $= 40 - 50$

$\therefore \quad 40 + \dfrac{10}{65} \left[115 - (a + 42) \right]$

$$a = \underline{\underline{34}}$$

$\therefore$ using eq ①

$$b = 80 - 34$$

$$b = \underline{\underline{46}}$$

* Mode

    Mode is the value of a variable occurring more number of times in the dataset i.e Highest frequency of the variable.

Case I   Raw Data

    The value that occurs the maximum number of times is the Mode.

Case II   Discrete frequency Distribution.

    The value of the variable with max frequency is the Mode

Case III   Continous frequency Distribution.

    The class with Max frequency is called Model Class then the Mode is obtained by the formula.

$$\text{Mode} = L + \left[\frac{f_M - f_1}{2f_M - f_1 - f_2} \right] \times h$$

where f_M = frequency of Model Class
             f_1 = Frequency of Pre-Model class
             f_2 = frequency of Post Model Class

* lets Solve Some Questions

Q. Calculate the Mode

     10 , 8 , 6 , 15 , 18 , 10 , 12

->   Mode = 10   { Most Repeated }

Q. Calculate the mode for the following Dataset.

size	12	20	24	27
frequency	5	8	13	4

->   Highest frequency = 13 , Mode = 24

Q. Calculate the Mode for the following Dataset.

Sales (lakhs)	58-60	60-62	62-64	64-66	66-68	68-70	70-72
no. of companies	12	18	25	30	10	3	2

->   Model Class = 64 - 66   { Highest frequency }

   f_M = 30
   f_1 = 25
   f_2 = 10

  using formula : $L + \left[\dfrac{f_M - f_1}{2f_M - f_1 - f_2} \right] \times h$

     =>   25.6

H.W

Q. Calculate the Mean, Median & Mode for the following Data.

Class	10-15	15-20	20-25	25-30	30-35	35-40	T
Frequency	11	20	35	20	8	6	100

Median x_i	12.5	17.5	22.5	27.5	32.5	37.5	
$f_i x_i$	137.5	350	787.5	550	260	225	2310

CF	11	31	66	86	94	100

→ Calculating all the Required values:

① Mean

using formula : $\dfrac{\Sigma f_i x_i}{\Sigma f_i}$ ⇒ $\dfrac{2310}{100}$ ⇒ $\boxed{231}$

② Median

Calculating $N/2 = 50$

Median Class $= 20 - 25$

Using formula : $L + \dfrac{h}{f}\left[\dfrac{N}{2} - CF\right]$

⇒ $20 + \dfrac{8}{7\,35}\left[50 - 31\right]$

$= \boxed{22.71}$

③ Mode

Model Class = 20 - 25

$$F_M = 35$$
$$f_1 = 20$$
$$f_2 = 20$$

using formula : $L + \left[\dfrac{f_m - f_1}{2 f_M - f_1 - f_2} \right] \times h$

$\Rightarrow$ $20 + \left[\dfrac{35 - 20}{70 - 20 - 20} \right] 5$

$\Rightarrow$ $\boxed{10}$