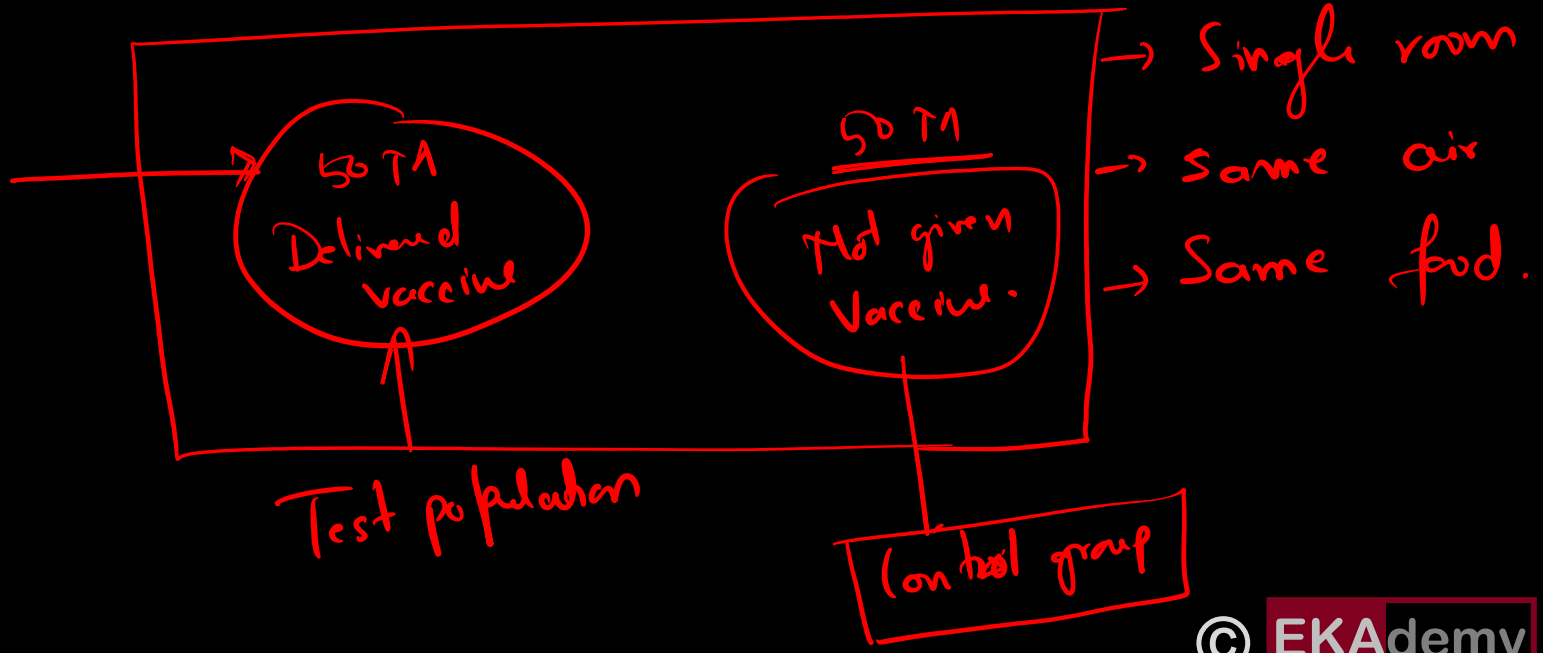
Statistics

Statistics

New Vaccine

1. 100 monkey's → Sample ~~popup~~ Population.



1 million health records in a hospital.

Statistics

→ { Pull out important/useful information from the given data.

Sample Population :→ Population chosen for study / trial / experiment.

- It should be representative of actual population
- It should be chosen randomly.
- Pull out data.

Data

↳ Study this data [apply statistical operations]

↓

Pull out important information from that data.

Data

- Collection of numbers / facts.

Two types data

① Primary data :→ When an investigator collects data himself/herself with a definite plan in his(her) mind, it is called primary data.

② Secondary data :→ Data which are not originally collected by the investigator but they are obtained from published or un-published sources, are known as secondary.

Organisation of data :

Raw data → first hand data that we collect from the field.

1. Serial order or alphabetical order
2. Ascending order
3. Descending order

eg:

Name	Marks
Sonu	72
Nikhil	89
Jhony	71
Priti	64

Raw Data.

Descending order

Name	Mark.
Nikhil	89
Sonu	72
Jhony	71
Priti	64

Alphabetical order

Name	Mark.
Jhony	71
Nikhil	89
Priti	64
Sonu	72

Ascending order

Name	Marks
Priti	64
Jhony	71
Sonu	72
Nikhil	89

Raw data.

39, 25, 5, 33, 19, 21, 12, 41, 12, 21, 19, 1, 10,
8, 12, 17, 19, 17, 17, 41, 40, 12, 41, 33, 19, 21, 33, 5,
1, 21.

Organise this data in ascending order.

⇒ 1, 1, 5, 5, 8, 10, 12, 12, 12, 12, 17, 17, 17, 19, 19, 19, 19,
21, 21, 21, 21, 25, 33, 33, 33, 39, 40, 41, 41, 41.

Organised data → ascending order ⇒ Array or Arrayed data.

Array or Arrayed data

↓

Table → Tabular representation of data.

Header ⇒

Marks	Tally marks	No. of students
1	||	2
5	||	2
8	||||	4
10	|	1
12	||||	4
17	||	3
19	||||	4
21	||||	4
25	|	1

→ | 41 | 3 |

Frequency

↓

No. of times an observation occurs in the given data.

⇒ Frequency distribution table

Raw data

No. of children in 20 families.

1, 1, 2, 3, 4, 3, 2, 1, 1, 4, 5, 2, 4, 2, 2, 1, 3, 3, 2, 5

Create a frequency distribution table for the given data.

Frequency distribution Table ⇒

No. of children	Tally	Frequency (No. of families)
1	𝍸	5
2	𝍸 I	6
3	IIII	4
4	III	3
5	II	2

Ages of 25 students:

→ 15, 16, 16, 14, 17, 17, 16, 15, 15, 16, 16, 17, 15

16, 16, 14, 16, 15, 14, 16, 15, 15, 15, 14, 15

Prepare a frequency distribution table for the given data set.

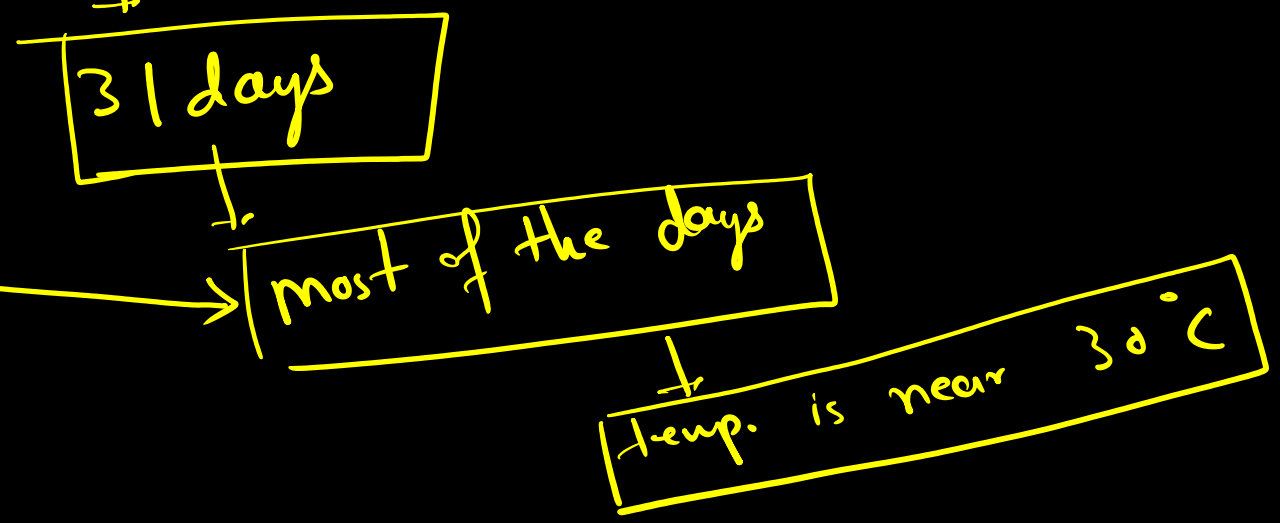
⇒

Age	Tally bars	frequency (No. of students)
14		4
15		8
16		10
17		3

Discrete Frequency distribution table.

Central Values

→ Average temp. of a January in K.L. is 30°C.



→ Average marks/score of class VII is 86.

[75, 89, 91, 84, 86, 85, 79, 87, 92, 84.

85

Central value

↓

around which most of the observation lie.

Depending upon data, there are various ways to calculate/find central value of the data.

(i) Arithmetic mean (mean)

(ii) Median

(iii) Mode

Mean (Average)

$$\text{Mean} = \frac{\text{Sum of all the observation}}{\text{No. of observation.}}$$

Eg: 86, 84, 81, 89, 87 — Marks of 5 student.

$$\text{mean} = \frac{86+84+81+89+87}{5} = \frac{427}{5} = 85.4$$

⇒ marks of most of the students is close to 85.4

$x_1, x_2, x_3, x_4, \ldots, x_n$ $\Rightarrow$ n observations

$$\text{mean} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$\bar{X}$ ↑

ex bar

$\Rightarrow$ Range : diff. between highest and lowest observation in the data set.

eg: → 2, 9, 15, 25, 7, 5, 4, 20, 91, 99

$$\bar{x} = \frac{2+9+15+25+7+5+4+20+91+99}{10} = \frac{272}{10}$$

$$= 27.7 \quad \times \text{ not a good (valid) value.} \qquad = \frac{277}{10}$$

Median : 2, 4, 5, 7, 9, 15, 20, 25, 91, 99

$$\frac{9+15}{2} = \frac{24}{2} = 12$$

12

1. If the mean of 6, 8, 5, (x) and 4 is 7. Find the value of x.

Sol.

$$\text{Mean} = 7$$

$$\frac{6+8+5+x+4}{5} = 7$$

$$\frac{23+x}{5} = 7$$

$$23 + x = 7 \times 5$$

$$23 + x = 35$$

$$x = 35 - 23 = 12$$

2. If mean of 6, 4, 7, p, 10 is 8, find the value of p.

Ans: 13

3. The mean of 10 numbers is 20. If 5 is subtracted from every number, what will be the new mean.

Sol. The 10 numbers are : $x_1, x_2, x_3, \cdots, x_{10}$

$$\text{Mean} = \frac{x_1 + x_2 + x_3 + \cdots + x_{10}}{10}$$

$$\therefore \quad \frac{x_1 + x_2 + x_3 + \cdots + x_{10}}{10} = 20$$

$$\Rightarrow \quad \boxed{x_1 + x_2 + x_3 + \cdots + x_{10} = 200} \quad \text{—①}$$

If 5 is subtracted from every no., the new numbers will be.

$$\Rightarrow (x_1 - 5), (x_2 - 5), (x_3 - 5), \cdots, (x_{10} - 5)$$

$$\text{New mean} = \frac{(x_1 - 5) + (x_2 - 5) + (x_3 - 5) + \cdots + (x_{10} - 5)}{10}$$

$$= \frac{(x_1 + x_2 + x_3 + \cdots + x_{10}) - 50}{10}$$

$$= \frac{200 - 50}{10} = 15$$

4. The mean of 16 numbers is 8. If 2 is added to every no., what will be the new mean?

(10)

5. If the mean of five observations x, $x+2$, $x+4$, $x+6$, $x+8$ is 11.

Find the mean of first 3 observation.

Sol

$$\frac{x + (x+2) + (x+4) + (x+6) + (x+8)}{5} = 11$$

$$x = 7$$

So, First 3 observations are : 7, 9, 11

Mean of 7, 9, 11 $= \frac{7+9+11}{3} = 9.$

5. The mean of 40 observation was 160. It was detected on re-checking that the value 165 was wrongly copied as 125 for computation of mean. Find the correct mean.

$$\text{Mean} = \frac{\text{Sum of all the observations}}{\text{No. of observations.}}$$

$$160 = \frac{\text{Sum of all the observation}}{40}$$

Sum of all the 40 observation $= 160 \times 40 = 6400$.

Incorrect sum of observation $= 6400$

Correct sum of the observation = Incorrect sum − Incorrect observation + Correct observation

$= 6400 - 125 + 165. = 6440$

$$\text{Correct mean} = \frac{\text{Correct sum of 40 observations}}{40}$$

$$= \frac{6440}{40}$$

$$= \frac{644}{4}$$

$$= \underline{\underline{161}}$$

Arithmetic mean of grouped data

S_1 →	160 cm.
S_2 →	120 cm.
S_3 →	80 cm.
S_4 →	200 cm.
S_5 →	80 cm.
S_6 →	160 cm
S_{100} →	

Mean =

Grouped data

x_i Height (cm) (observation)	f_i No. of student (frequency)	$x_i \cdot f_i$
x_1 80	f_1 14	14×80
x_2 110	f_2 16	16×110
x_3 140	f_3 40	40×140
x_4 160	f_4 25	25×160
x_5 165	f_5 15	15×165
$n =$ 100		$\sum x_i f_i$

$$\text{Mean} = \frac{(x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_n f_n)}{(f_1 + f_2 + f_3 + \cdots + f_n)} = \frac{\sum x_i f_i}{\sum f_i}$$

$\sum$ => Sigma.

Sum of.

$$\text{Mean} = \frac{\sum x_i f_i}{\sum f_i}$$

Find the mean of given data.

Age (x_i)	frequency (f_i)	$x_i f_i$
14	1	14 (14×1)
15	5	75 (15×5)
16	12	192 (16×12)
17	7	119 (17×7)
	$\sum f_i = 25$	$\sum x_i f_i = 400$

$$\text{Mean} = \frac{\sum x_i f_i}{\sum f_i} = \frac{400}{25} = 16 \text{ years}$$

T/Marks in a unit test.

$\Rightarrow$ [4, 6, 7, 5, 3, 5, 4, 5, 8, 1, 2, 5, 1, 9, 6, 7, 8, 4, 6, 7] = 108

Organise this data in tabular form. Find the arithmetic mean of marks. Also, find the highest marks, lowest mark and range of the data.

x_i (marks)	f_i (no. of students)	$x_i \cdot f_i$
1	1	1
2	1	2
3	1	3
4	3	12
5	4	20
6	4	24
7	3	21
8	2	16
9	1	9

$\sum f_i = 20$ $\quad \sum x_i f_i = 108$

$$\text{Mean} = \frac{\sum x_i f_i}{\sum f_i}$$

$$\text{mean} = \frac{108}{20}$$

$$= 5.4$$

Range: $9 - 1 = 8$

Data: Find the mean of given data:

x_i	4	6	9	10	15
f_i	5	10	10	7	8

Sol.

x_i	f_i	$x_i \cdot f_i$
4	5	20
6	10	60
9	10	90
10	7	70
15	8	120

$\sum f_i = 40$ $\sum x_i f_i = 360$

$$\text{mean} = \frac{\sum x_i f_i}{\sum f_i}$$

$$\text{mean} = \frac{360}{40} = 9$$

x_i	2	4	6	10	$p+5$
f_i	3	2	3	1	2

If the mean of the above data is 6, find the value of p.

x_i	f_i	$x_i f_i$
2	3	6
4	2	8
6	3	18
10	1	10
$(p+5)$	2	$2p+10$

$\Sigma f_i = 11$ $\Sigma x_i f_i = 2p+52$

$$\text{Mean} = \frac{\Sigma x_i f_i}{\Sigma f_i}$$

$$\text{Mean} = \frac{2p+52}{11}$$

$$\frac{2p+52}{11} = 6$$

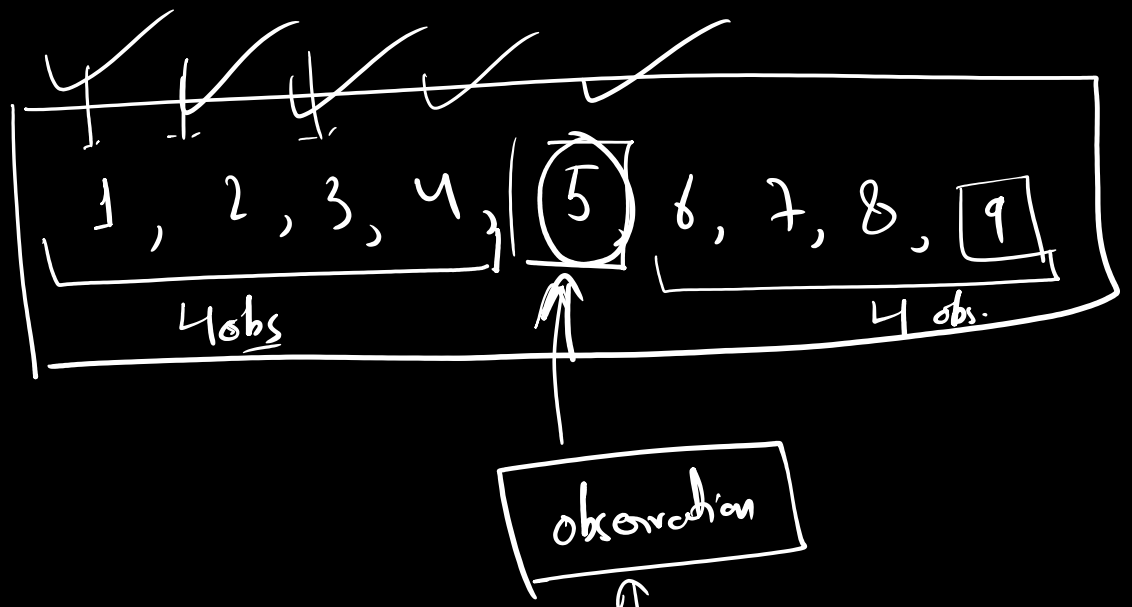
$$2p+52 = 66$$

$$2p = 14$$

$$p = 7$$

Median

Data set



odd

$\frac{9 \text{ observ} + 1}{2}$

$\frac{10}{2} = 5^{th}$ obs.

no. of observation = n

n = odd

median =

$\Rightarrow$ 11, 12, 13, 14, (15), 16, 17, 18, 19,

↑ 5th ob.

n = odd.

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

n = no. of observations

n = 9

$$= \left(\frac{9+1}{2}\right)^{th} \text{ observation}$$

$$\text{median} = 5^{th} \text{ observation}$$

median = 15

n = even

11, 12, 13, 14, 15, 16, 17, 18, 19, 20

(14 and 15 are the middle observations; 4 observations on each side)

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

$n = 10$

$$\text{median} = \frac{\left(\frac{10}{2}\right)^{th} \text{ observation} + \left(\frac{10}{2}+1\right)^{th} \text{ observation}}{2}$$

$$= \frac{5^{th} \text{ observation} + 6^{th} \text{ obs}}{2}$$

$$= \frac{15+16}{2} = 15.5$$

Ex. Find the median of the values: 37, 31, 42, 43, 46, 25, 39, 45, 32.

Sol: (1) Arrange in ascending.

25, 31, 32, 37, 39, 42, 43, 45, 46

$n = 9$

$$\text{median} = \left(\frac{9+1}{2}\right)^{\text{th}} \text{ observation}$$

$$= \left(\frac{10}{2}\right)^{\text{th}} \text{ obs.}$$

$$= 5^{\text{th}} \text{ obsn.}$$

$$= \underline{39}$$

Find the median :

25, 34, 31, 23, 22, 26, 35, 28, 20, 32.

$$= \frac{5^{th} \text{ obs} + 6^{th} \text{ ob.}}{2}$$

$$= \frac{26 + 28}{2}$$

median = 27.

Ex. The median of the observations 11, 12, 14, 18, $x+2$, $x+4$, 30, 32, 35, 41 is 24. Find the value of x.

$n = 10$

$$\text{median} = \frac{\left(\frac{10}{2}\right)^{th} \text{ obs} + \left(\frac{10}{2}+1\right)^{th} \text{ obs.}}{2}$$

$$= \frac{5^{th} \text{ obs} + 6^{th} \text{ obs.}}{2}$$

$$24 = \frac{(x+2) + (x+4)}{2}$$ Linear eq.

$$24 \times 2 = [(x+2) + (x+4)] \times 2$$

$$48 = (x+2) + (x+4)$$

$$48 = 2x + 6$$

$$2x = 48 - 6$$

$$2x = 42$$

$$x = 21$$

Find the median of the data: 19, 25, 59, 48, 35, 31, 30, 32, 51.

If 25 is replaced by 52, what will be the new median.

Mode

↳ observation which occurs most frequency.

Empirical formula.

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

⇒ There can be more than one mode in the given data set.

[38, 42, 35, 37, 45, 50, 32, 43, 43, 40, 36, 38, 43, 36, 47]

Find the mode and median of this data.

Array: 32, 35, 36, 37, 38, 38, 38, 40, 42, 43, 43, 43, 45, 47, 50

(3 time) (3 time.)

mode

Mode: 2 modes are 38 and 43

median:

12, 14, 12, 16, 18, 15, 14, 18, 19, 12, 14, 15, 16, 15, 16, 16, 15, 17, 13, 16, 16, 15, 15, 13, 18, 17, 18, 14, 15, 13, 15, 14

Mode and ~~median~~ mean.

mode: 15

x_i	f_i	$x_i f_i$
12	3	36
13	4	52
14	5	70
15	10	150
16	6	96
17	2	34
18	1	18
19	1	19

$\sum x_i f_i =$

$$\text{mean} = \frac{\sum x_i f_i}{\sum f_i}$$

H. W.

End of the Chapter