

probability and Statistics

unit 1

Basic Statistics

measures of central tendency.

measures of dispersion.

Coef. of Variation.

movements

Skewness : How the distribution is deviated from symme.

Kurtosis

Unit 2 random variables (rv)

- Introduction to RV
- Distribution function for discrete and contin. rv.
- Commulative distribution fun
- Joint distribution
- Joint density fun.
- mathematical expectation
- variance

unit 3

correlation and regression.

→ Least Square-approximation: fitting of straight line

Second degree parabola and exponential fit.

→ Correlation, rank correlation and linear regression.

unit 4

probability distribution

→ standard discrete distribution

→ basic probability

(a) binomial distribution

→ Conditional probability.

(b) Poisson distribution

→ Independent events

(c) exponential dis.

→ Bayes theorem

(d) Geometric dis.

→ standard cont'n distribution:

a) Normal distribution.

units

estimation and testing of hypothesis.

- a) Sampling and estimation of parameters
- b) Standard error and Statistical hypoth.
- c) testing a hypothesis
- d) One and tailed test of hypo.
- e) type one and type two errors
- f) ~~table~~ of significance Level
- g) chisquare test
- h) Students t-tests, F-test

Introduction to Statistics:

- 1) Variable: → A quantity which can vary from one individual to another is called Variable.
- two types:
- a) discrete (discontⁿ):
ex: no. of students, +ve integer.
no. of trees
 - b) Contⁿ Variable
ex: height, weight ... etc.

- 2) frequency distribution:
consider marks obtained by 60 students of Sy. in a test.

↳ (we have to arrange data in $\downarrow$ or $\uparrow$ order) according to the roll no. So to study the random data is required to arrange the data in asc or $\hat{a}$ des order.

So the conclusion will be: min / max marks are observed but to arrange it in more compact form, it is presented in a tabular form.

Three types of distributions are: ① IP (individual data) ②

1) IP: (freq. are not given) i.e. $x_1, x_2, x_3, \dots, x_n$
at most ($n=150$) can be.

2) DP: discrete data (freq. are provided)
ex, $x: x_1, x_2, x_3, \dots, x_n$
 $f: f_1, f_2, f_3, \dots, f_n$

3) GD (CD)

Grouped frequency distribution is of the form

ex: class interval (CI) : $a_1 - a_2$ $a_2 - a_3$ $a_3 - a_4$ -
freq. : f_1 f_2 f_3 - -

* inclusive class interval: $[a, b]$

* exclusive class interval: $[a, b)$

ex: 99.5 $\rightarrow$ previous interval modify करेंगे.

4) Location of central tendency:

after collecting a data and arranging it in a proper order in the form of freq dis. (appropriate freq. distribution) the next task is to study this data to find valid conclusion, it is req. to find central tendency known as:

- a) AM
- b) mode
- c) medium

1) A.M ($\bar{x}$)

a) ID: $\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$

b) D.D

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

where $\sum f_i = N$.

c) G.D / C.D

$N = \text{no. of observation}$

CT	mid-value	f	$u = \frac{x-A}{h}$	$f_i u_i$
----	-----------	---	---------------------	-----------

mid-value $\rightarrow x$

therefore

A - assumed mean

h - class width

$$\bar{x} = A + h \frac{\sum f_i u_i}{\sum f_i}$$

$$\text{i.e. } \bar{x} = A + h \cdot \frac{\sum f_i u_i}{N}$$

Que) find A.M for the following data

X	f	$x_i f_i$
5	3	15
9	6	54
12	8	96
15	8	120
20	9	180
24	10	240
30	8	240
35	7	245
42	6	252
49	2	98
	67	1540

$$\bar{x} = \frac{1540}{67} = 22.98$$

any can choose betw: 55, 45

classmate

Date

Page

que) find AM for the following data

C.I marks	f No. of Students	Σ mid value	$u = \frac{x-A}{h}$	fiu
0-10	8	5	-4	-32
10-20	20	15	-3	-60
20-30	14	25	-2	-28
30-40	16	35	-1	-16
40-50	20	<u>45</u>	0	0
50-60	25	55	1	25
60-70	13	65	2	26
70-80	10	75	3	30
80-90	5	85	4	20
90-100	<u>2</u>	95	5	10
	133			-25

$$\bar{x} = \frac{45 + 10 \times (-25)}{133}$$

$$= 45 + -1.88$$

$$= \boxed{43.12}$$

for

Now

$$\boxed{A=55}$$

$$\bar{x} = \frac{55 - 10 \times (-25)}{133}$$

$$= 55 - 1.88$$

$$= \boxed{53.12}$$

2) median

→ median is a central value of the variable when the values are arranged in ascending or descending order

→ media divides the distribution into two equal parts, it is also called potential average.

formulas for

① ID

• In the case of ungrouped data:

Case 1 If no. of observation is odd.
↳ 'n'

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

e.g : 1, 9, 5, 3, 11, 24, 27

→ median ?

asc: 1, 3, 5, 9, 11, 24, 27

Now

median: 9

4th observation = 9 is the median for the given data.

Case 2

If n is even then median is average of two middle terms.

ex: 1, 9, 5, 3, 16, 11, 24, 27

asc. order: 1, 3, 5, 9, 11, 16, 24, 27

Now central values: $\frac{20}{2} = \boxed{10}$

The median of the even terms may or may not belong the given data.

⑥ DD
→ In case of discrete frequency distribution construct commulative frequency column then calculate $\sum f_i = N = \text{total frequency}$ then median is, value just $\geq \frac{N}{2}$ in the commulative freq. column the corresponding value of x .

[corresponding value of x is the median]

x :	1	2	3	4	5	6	7	8	9
f :	8	10	11	16	20	25	15	9	6
cf :	8	18	29	45	65	90	105	114	120

$$\text{i.e. } \frac{N}{2} = \frac{120}{2} = 60$$

the value which is greater $\geq \frac{N}{2} = 65$

Here we observed that commulative frequency just greater than equal to 60 is observed as 65 and hence the corresponding value of $x = 5$ is the median.

③ GD/CD

for grouped frequency distribution median is

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

where

l = lower limit of median class

f = freq. of the median class

h = width of the median class

c = cumulative freq. of the class preceding to the median class.

3 ④ MODE:

mode is the value which occurs most frequently in the set of observation.

i.e. the value of variate corresponding to the max. freq.

Ex: a) ID: mode may or may not exist.

b) DD: In case of discrete data mode is the value of X , corresponding to the max. freq.

Same ex:

X :	1	2	3	4	5	6	7	8	9
	8	10	11	16	20	25	15	9	6

i.e. mode = 6

Here we observed that $f_{\max} = 25$

i.e. corresponding value of $X = 6$

c) GD :

$$\text{mode} = l + h \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right)$$

Where l : Lower limit of modal class h : width of modal class f_1 : frequency of modal class f_0 : frequency of the class preceding to the modal class. f_2 : frequency of the class succeeding to the modal class.

Ques) Find mean, mode, median for the following data.

mid-val.

1)	C.I	x	f_i	u_i	$f_i u_i$
	0-10	5	4	-3	-12
	10-20	15	7	-2	-14
	20-30	25	8	-1	-8
	30-40	35	12	0	0
	40-50	45	1	1	1
	50-60	55	2	2	4
	60-70	65	3	3	9

$$\Sigma f_i = 37$$

$$\Sigma f_i u_i = -20$$

A. 35

$$\text{mode} = 30 + 10 \frac{(12-8)}{2 \cdot 12 - 8 - 1}$$

$$= 30 + 10 \frac{4}{15}$$

$$= 30 + 2.66$$

$$32.66$$

$$= 32.66$$

mean: $\bar{X} = 29.59$

Median: $l + \frac{\frac{n}{2} - c}{f} \left(\frac{N}{2} - c \right)$

$$20 + \frac{10}{8} (17.5 - 11) = \boxed{29.315}$$

How we observed that AM for all three

series is same.

but in series A observations does not

deviate from AM as in the deviation.

To series B one observation is represented

by AM and other observations are slightly

deviated from AM as compared with series

To series C not a single observation is

represented by AM and there is a wide

range of deviation as compared with series

To series D central tendency is not

to confirm the data and there is a wide

range of deviation as compared with series