Lecture-2
10th Jan. 2024

A quantity which can vary from one individual to another is called variable or variate

PRIMARY DATA: original & first hand information

SECONDARY DATA: ~~copy & second hand~~ collected from records or already available

GEOGRAPHICAL CLASSIFICATION

→ Data is ~~not~~ classified as per geogra-phical region or place.

→ Eg. Production of wheat in diff. ~~regions~~ countries.

QUALITATIVE CLASSIFICATION

→ Data is classified on the basis of quality.

TYPES OF VARIABLE

Qualitative

→ ~~is~~ cannot be measured in terms of magnitude & can be expressed in terms of qua-lity or kind

Eg. Nationality
Occupation
Religion
Literacy
Beauty

Quantative

→ can be measured numerically on the scale.

Eg. Age
Height
Speed
Weight

Discrete

A variable which takes only finite values with-in a certain range

Eg: No. of children in a family
No. of accidents per week

Continuous

A variable which take any numeri-cal value within a certain rang ~~(is called)~~

Eg. Age, Heig-t, weight, distance

QUATITATIVE CLASSIFICATION AND FREQUENCY DISTRIBUTION

Let the marks obtained by 40 students of AI out of 50 marks

5, 6, 7, 8, 11, 15, 20, 8, 11, 25, 30, 15, 17, 11,
6, 22, 25, 20, 22, 15, 30, 32, 32, 8, 20, 25, 22,
22, 35, 37, 40, 20, 11, 25, 10, 20, 10, 15, 35, 42

Marks	Sample Space	Tally Bar	Frequency	cf → greater than
5-10	5	𝍸 II	7	40
10-15	5	𝍸 𝍸 I	6	33
15-20	5	𝍸	5	27
20-25	5	𝍸 𝍸 IIII	9	22
25-30	5	IIII	4	13
30-35	5	IIII	4	9
35-40	5	III	3	5
40-45	5	II	2	2
~~45~~				

Sample space must be equal in all groups

c.f
7
7+6
7+6+5
7+6+5+9
7+6+5+9+4
.
.
.
.

Summary of Dec-4

- Definition of C.F
- Concept of Exclusive & Inclusive class interval
- How to convert Inclusive class into Exclusive class interval

CUMMULATIVE FREQUENCY

Some times one may want to know less than or greater than a particular value.

This is done by finding cummu-lative frequency.

The CF is (sometimes called 'less than CF) corresponding to a class to the class is the sum of the frequencies of that class.

A table displaying the class interval and CF is called as CF table

EXCLUSIVE AND INCLUSIVE CLASS INTERVALS

The class intervals open to the right are called Exclu-sive as the exclude the upper limit of the class

[a, b) [a → lower limit, b → upper limit]

Eg:

Income (in ₹)	No. of person
250 - 500	75
500 - 750	70
750 - 1000	52
1000 - 1250	33

↑ exclusive class interval

In ECI, the upper limit of upper class is the lower limit of next class

If the ₹ income of a person is 750 the he is included in class 750-1000.

The class interval is closed to the right are called inclusive since they include the upper limit of the class

[a,b] [a → LL, b → UL]

compare class width's before adj & after adj

Eg:

Income (in ₹)	No. of person
250 - 499	60
500 - 749	43
750 - 999	25

4000

However, for further statistical analysis is desirable that class interval is exclusive.

Jump

To convert inclusive class int. into ECI:

① Find the diff. b/w LL of second class to the UL of the first class.

② Subtract the half of this diff. from all the LL and add to all the U.L.

③ The adjustment factor in the above eg. is: $\frac{500 - 499}{2} = 0.5$

④ Now, from all the lower limits → add 0.5 and from all the upper limit → subtract 0.5

16/01/2024

Lecture 5 Summary-

- Histogram formation
- Frequency Polygon formation

GRAPHICAL REPRESENTATION OF DATA

The Grouped frequency distribution provides a better idea of the data as compared to the ungrouped frequency dist.

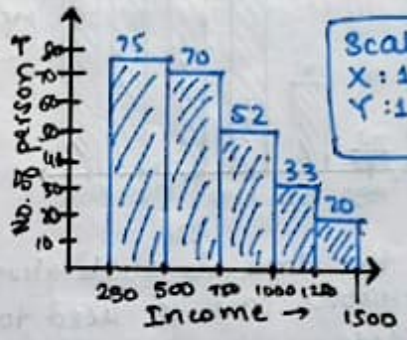
Further, if the distribution is represented by graphs a more clear visual impression about the data is obtained as graphs are the ~~grouped~~ good visual aid.

various imp. types of graphs :-

→ Histogram
→ Bar Charts
→ Frequency curve
→ Frequency polygon

① HISTOGRAM

INCOME (in ₹)	No. of Persons
250 - 500	75
500 - 750	70
750 - 1000	52
1000 - 1250	33
1250 - 1500	20



Scale:
X : 1 unit → 250
Y : 1 unit → 10 person

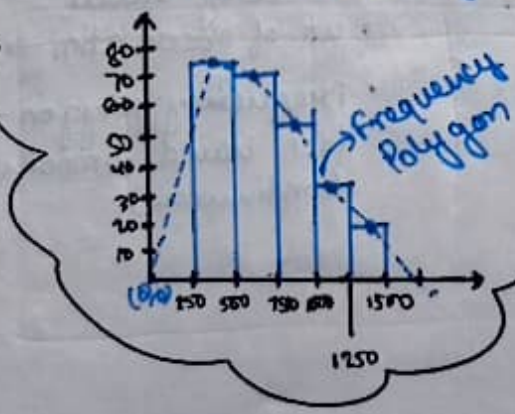
To draw the histogram

1. Mark all the class interval on X-axis
2. Mark frequencies along Y-axis
3. It is not necessary to take same scale on both axis. You can take different scales.
4. Construct rectangles for each class interval having the height equal to the corresponding frequency.

② FREQUENCY POLYGON (FP)

For a grouped frequency distribution with equal class interval a FP is obtained by:-

1. Joining the middle points of the top of the rectangles of the histogram by means of dotted straight lines.
2. To complete the polygon the mid pt. at each ~~end~~ end are joined to the just lower & higher mid pts of zero frequency (origin)



Lecture-6 | Summary

17th Jan! 24

- Bar Chart
  - → Simple Bar Diagram
  - → Multiple Bar Diagram
- → Frequency Curve

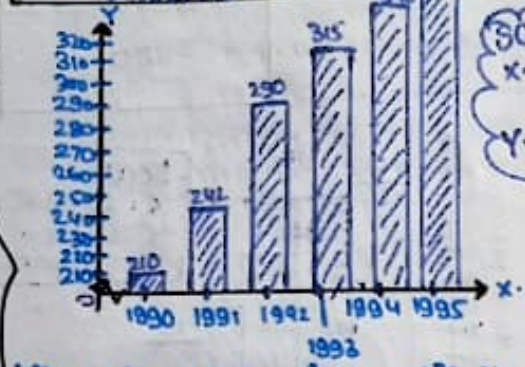
BAR CHART

These are also called as bar graphs. Various types of bar graphs:-

→ Simple Bar Diagram

E.G.

YEAR	1990	1991	1992	1993	1994	1995
NO. OF STUDENTS	210	242	290	315	340	355



SCALE
x-axis: 1 unit → 1 year
Y-axis: 1 unit → 10 people

When the data is qualitative or categorical bar graphs can be used to represent the data.

(i) A bar graph can be drawn horizontally or vertically.

NOTE: The height of the bar represent the frequency of data.

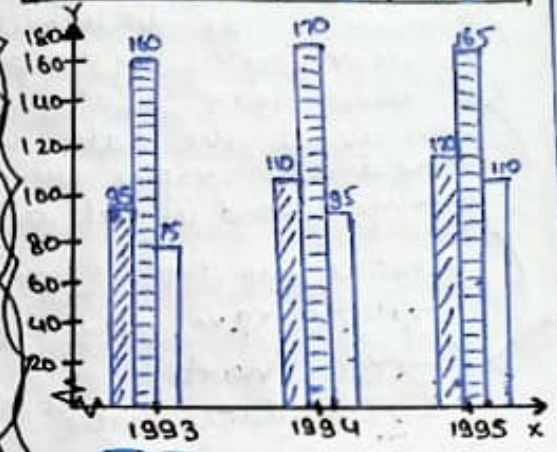
(ii) The distance b/w two bars should be constant

(iii) The width of each bar must be same.

→ Multiple Bar Diagram

E.G.

YEARS	1993	1994	1995
ARTS	95	110	120
SCIENCE	160	170	165
COMMERCE	75	95	110



SCALE
X-axis: 1 unit → 1 year
Y-axis: 1 unit → 20 students

▨ = Arts
▤ = Science
□ = Commerce

→ Frequency Curve

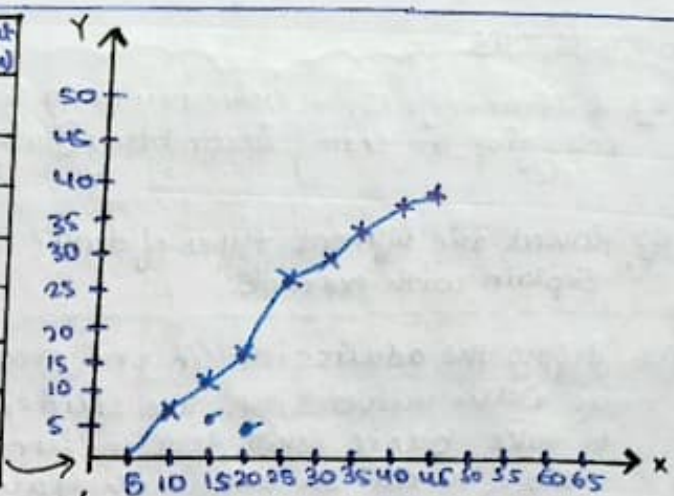
- Grouped frequency distribution
- width of class int. → very small
- no. of observation → Increases
- Frequency polygon → free hand smooth curve (continuous)

Lecture-7 (Summary)

19th Jan 24

- Less than ogive curve plotting
- More than ogive curve plotting

CI	Tally	Freq	CF (less than)	CF (greater than)
5-10	𝍸 II	7	7	40
10-15	𝍸 I	6	13	33
15-20	𝍸	5	18	27
20-25	𝍸 IIII	9	27	22
25-30	IIII	4	31	13
30-35	IIII	4	35	9
35-40	III	3	38	5
40-45	II	2	40	2

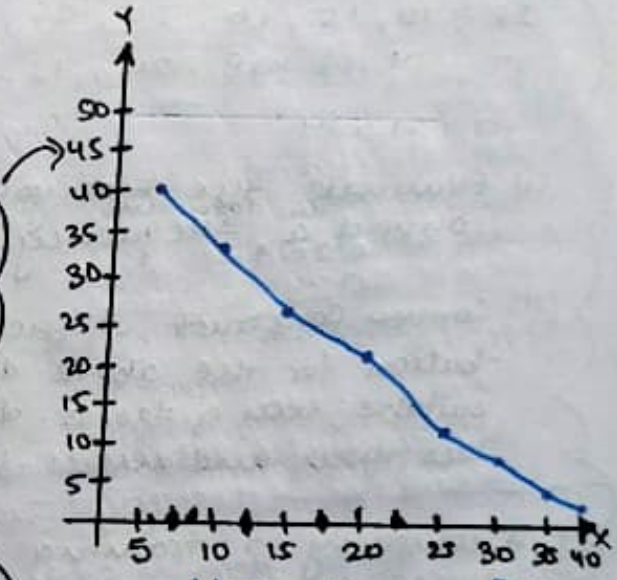


Less than ogive

There are two types of curve
(i) less than ogive
(ii) More than ogive

For less than ogive:-
→ we plot the points corresponding to upper limit on x-axis and (less than) CF on Y axis. If a free hand smooth curve and joining the points we get less than ogive.

For More than ogive:-
→ we plot the points corresponding to lower limits on X-axis and (more than) CF on Y-axis. A free hand smooth curve joining the points we get more than ogive



More than ogive

{Lecture-8

23rd January 2024 Summary

- Important type of question

IMP QUES TYPE

SCALE - MEASUREMENT OF SCALE

Q1. Difference b/w - Primary Data & Secondary Data (also give an example) → Methods for collecting Primary Data & sec. Data

Q2. What are different types of data? Explain with example

Q3. Before the admission in phd program in Delhi university. The student have to take basic skill test in fundamentals of math. In one such example, 30 student appeared in the exam. They are marked out of 30 points.

15, 12, 15, 22, 28, 30, 19, 25, 24,
28, 10, 15, 16, 20, 26, 22, 18, 20,
27, 14, 12, 19, 21, 16, 19, 30, 13,
10, 21, 24

(i) Eliminate the steps you take to convert a raw data into a frequency distribution

(ii) ~~Convert~~ Construct a frequency distribution for the above data with a suitable no. of classes. Also construct less than and more than ogive.

Q4. A company a training program after the first week, the company officers joining the training program. The score out of 100 of 80 employees:-

32, 36, 31, 67, 65, 42, 39, 56, 78, 61,
34, 78, 75, 78, 61, 30, 65, 45, 48, 78
43, 75, 64, 73, 87, 41, 56, 71, 81, 85

(i) convert the data into frequency distribution by taking a suitable width.

(ii) construct a histogram & frequency polygon diagram.

Lecture-9 Summary

24th January '24

- Scale
- Mean
- Measures of central Tendancy

SCALE

1. Nominal Scale: on basis of quality
2. Ordinal Scale: on basis of order
3. Interval Scale: " " " interval

MEASURE OF CENTRAL TENDANCY

- Single value that attempt to describe a set of data by identifying the central position within that set of data.
- As such of central tendancy are sometimes called measures of central location
- It is a way of summarising the data in a form of a typical or representative values
- Four types of mean:-
  - → Arithmatic mean
  - → Geometric mean
  - → Harmonic mean

CENTRAL TENDANCY OF DATA

1. Mean

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

→ $\Sigma (x - m) = 0$

→ $\Sigma (x - m)^2$ → minimum

→ mean = μ

Mean of sample = $\bar{X}$

→ $\bar{x} = \frac{\Sigma fx}{N}$ [Mean of grouped Data]

→ use mean when:-
- data are interval or ratio scale
- data are not skewed
- because it is sensitive to every score

9th Jan '24

Lecture -10 Summary

- Basic formulae
- Mean → merit → ways to calculate → Direct Method, Step deviation method
- Median → merits & Demerit

Variance = $\frac{\Sigma(x_i - \bar{x}_i)^2}{n}$ [in form of mean]

Coeff. of variance = $\frac{\text{Mean}}{\text{S.D.}}$

Variance = $\frac{\Sigma(x_i - \text{Median})}{\text{Mean Deviation}}$ [in form of median]

Q) x_i: 3.2, 5.8, 7.9, 4.5

f_i: x, $x+2$, $x-3$, $x+6$

Find x if AM = 4.876 ($\bar{x}$)

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

Substitute & get the answer

$4.876(4x+15) = 2.1x + 14.9$

$x = 30.717$

MEDIAN

→ Positional number which divides the given series into two parts.

ANSWER: Median is 3rd position

MERITS

- Not affected by extreme obs.
- can be computed with open & closed intervals.

DEMERT:

- In case of even number of obs. for grouped data, median cannot be determined exactly
- Not based on each and every items of distribution
- cannot be used to compute the median of combined group.
- Median is less stable because it depends on its position.

Mean

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$ (Arithmetic Mean)

DEMERIT:

1. AM is affected by extreme observations
2. AM can not be used in case of open & closed classes.
3. AM cannot be used to calculate when dealing with qualitative information
4. AM can be used if any one observation is missing.

DIRECT METHOD

$x = \frac{L+U}{2}$ [Mid pt. of CI]

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

STEP DEVIATION

$\text{Mean} = A + h\frac{\Sigma fd}{\Sigma f}$

$d = \frac{x + A}{h}$ → assumed mean; h → class size

Middle value of CI

Marks	No. of St.	Mid pt. (x_i)	$d = \frac{x-A}{h}$	fd
0-10	6	5	-3	-18
10-20	5	15	-2	-10
20-30	8	25	-1	-8
30-40	29	[A = 35]	0	0
40-50	7	45	1	7
50-60	6	55	2	12
60-70	3	65	3	9
	$\Sigma f = 60$			$\Sigma fd = -8$

$h = 10$

$A = 35$

$x = \frac{L+U}{2}$

$\bar{X} = A + h\frac{\Sigma fd}{\Sigma f}$

$\bar{X} = 35 + 10\left(\frac{-8}{60}\right)$

$\bar{X} = 33.4$

• Question Solving

Question

Calculate the mean & std. deviation for the following table given age distribution of 542 members

Age Group	f
20 - 30	3
30 - 40	61
40 - 50	132
50 - 60	153
60 - 70	140
70 - 80	51
80 - 90	2

variance formula

$$\sigma^2 = h^2 \left[\frac{1}{N}(\Sigma f d^2) - \left[\frac{1}{N}(\Sigma f d) \right]^2 \right]$$

$N = \Sigma f$

do $h = 10$; $A = 55$

Age	f	x	$d = \frac{x-A}{h}$	fd	fx_i	fd^2
20-30	3	25	-30	-90	75	27
30-40	61	35	-20	-122	2135	244
40-50	132	45	-10	-132	5940	132
50-60	153	55	0	0	8415	0
60-70	140	65	10	140	9100	140
70-80	51	75	20	102	3825	204
80-90	2	85	30	60	170	18
	sum = 542			sum = -15	sum = 29660	Σ = 765

(i) Direct Method –

$$\bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{29660}{542} = 54.72$$

(ii) Step deviation

$$\bar{X} = A + h\left(\frac{\Sigma fd}{\Sigma f}\right)$$

$$= 55 + 10\left(\frac{-15}{542}\right)$$

$$= 55 - 0.27 = 54.73$$

(iii) Variance =

$$100\left[\frac{1}{542}(765) + \left(\frac{1}{542} \times -15\right)^2\right]$$

$$= 100\left[1.411 + 0.0007\right]$$

$$= 141.17$$

Standard Deviation = $\sqrt{\text{Variance}}$

$$= \sqrt{141.17}$$

$$= 11.88 \text{ Years}$$

31st January 2024

Lecture -12 Summary

- Question on mean
- Median
- Question on median of continuous class series

(1) Calculate the average marks of the following students

MARKS	NO. OF STUDENTS
0 -10	5
10-20	12
20-30	15
30-40	25
40-50	8
50-60	3
60-70	2

(2)

CI	f	x	$d=\frac{x-A}{h}$	fd	fd^2	fixi
0-10	5	5	-3	-15	45	25
10-20	12	15	-2	-24	48	180
20-30	15	25	-1	-15	15	375
30-40	25	35	0	0	0	875
40-50	8	45	1	8	8	360
50-60	3	55	2	6	12	165
60-70	2	65	3	6	18	130
	$\Sigma f = 70$			$\Sigma fd = -34$	$\Sigma fd^2 = 146$	$\Sigma fixi = 2110$

Step Deviation

1) Direct Method :-

$$\bar{x} = A + h\left(\frac{\Sigma fd}{\Sigma f}\right)$$

$$= 35 + 10\left(\frac{-34}{70}\right)$$

$$= 30.14 \text{ Marks}$$

2) Direct Method

$$\bar{x} = \frac{\Sigma fixi}{\Sigma fi}$$

$$= \frac{2110}{70}$$

$$= 30.14 \text{ Marks}$$

(3) MEDIAN

(i) For discrete series

$\left(\frac{n+1}{2}\right)^{th}$; when 'n' is odd

$\frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n+1}{2}\right)^{th}}{2}$; n ∈ even

(ii) For continuous series

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

l = lower limit of median class

h = class size of median class

f = frequency of median class

N = Total freq. Σfi

cf = cumulative frequency just preceding the median class

3) Variance

$$\sigma^2 = h^2\left[\frac{1}{N}\Sigma fd^2 - \left(\frac{1}{N}\Sigma fd\right)^2\right]$$

$$\sigma^2 = 100\left[\frac{1}{70} \times 146 - \left(\frac{-34}{70}\right)^2\right]$$

$$\sigma^2 = 100\left[2.08 - 0.23\right]$$

$$\sigma^2 = 185$$

4) Step deviation (σ) $= \sqrt{\sigma^2}$

$= \sqrt{185} = 13.6$

(4) Find the median of data

Age	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Person	8	12	10	8	3	2	7

AGE	f	cf
0-10	8	8
10-20	12	20
20-30	10	30
30-40	8	38
40-50	3	41
50-60	2	43
60-70	7	50
	N = 50	

Median class

$\frac{N}{2} = \frac{50}{2} = 25$

25 से just greater than cf.

To find median class

f = 10; l = 20; h = 10; N = 50; cf = 20

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

$$= 20 + \frac{10}{10}\left[\frac{50}{2} - 20\right] = 20 + 5 = 25 \text{ Ans}$$

2nd February 2024 — Lecture-18 Summ.

- Mode
- Question Practice

MODE

$$\text{Mode} = l + h\,\frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)}$$ (for continuous case)

$2f_1 - f_0 - f_2 \neq 0$

l = lower limit of modal class

h = class size of modal class

f_0 = frequency of class preceding the modal class

f_1 = freq. of modal class

f_2 = freq. of class just exceeding the modal class

for verification of Mode

Mode = 3Median − 2Mean

never −ve

AGE	PERSON
0-10	8 → f_0
10-20	12 → f_1
20-30	10 → f_2
30-40	8
40-50	3
50-60	2

Modal class = highest freq.

10-20

$l = 10$

$h = 10$

$$\text{Mode} = 10 + 10\left(\frac{12-8}{(2\times12)-8-10}\right)$$

$$= 10 + 10\left(\frac{4}{24-18}\right)$$

$$= 10 + \frac{40}{6}$$

$$= 10 + 6.67$$

$$= 16.67$$

Find the mode of the following:-

AGE	PERSON
0-10	8
10-20	12
20-30	10
30-40	8
40-50	3
50-60	2

Mode can be more than 1 in number

Lecture-14

5th Feb. 2024

- Method of grouping

Question

From the following data, regarding weight of 60 student of a class. Find mode.

weight =	50	51	52	53	54	55	56	57	58	59	60
No. of st. =	2	4	5	6	8	5	4	7	11	5	3

use 6 column to answer

① In column (1) write the original frequency

② In column (2), combine the frequencies two-by-two starting from top.

③ In column (3), combine the frequencies two-by-two starting from second class.

④ In column (4), sum up combine the frequencies three-by-three starting from top.

⑤ In column (5), combine three frequencies three-by-three starting from second.

⑥ In column (6), combine the freq. three-by-three starting from third.

weight	(1) f	(2)	(3)	(4)	(5)	(6)
50	2	6				
51	4		9	11		
52	5	11			15	
53	6		14			19
54	8	13		19		
55	5		9		17	
56	4	11				16
57	7		18	22		
58	11	16			23	
59	5		8			19
60	3					

only applied on irregular frequency data

irregular data

kabhi increase kr dia kabhi decrease kr dia

Column	Max. Freq.	class corresponding to max freq.			
(1)	11			58	
(2)	16			58	59
(3)	18		57	58	
(4)	22	56	57	58	
(5)	23		57	58	59
(6)	19	52 53 54	57	58	59 → 3 60 → 1
			4	6	

If the frequencies are ascending then apply the normal formula of mode

mode = 58

6th February 2024

Lecture - 15 Sum.

The median & mode of the following salary distribution are 3350 & 3400 resp. Sum of all freq. = 230.

SALARY	No. of EMPLOYEE
0-1000	4
1000-2000	16
2000-3000	f_3
3000-4000	f_4
4000-5000	f_5
5000-6000	6
6000-7000	4

①

SALARY	f	CF (less)
0-1000	4	4
1000-2000	16	20
2000-3000	f_3	$20+f_3$
3000-4000 (Median class)	f_4	$20+f_3+f_4$ (Modal class)
4000-5000	f_5	$20+f_3+f_4+f_5$
5000-6000	6	$26+f_3+f_4+f_5$
6000-7000	4	$30+f_3+f_4+f_5$

② $N = 230 = 30 + f_3 + f_4 + f_5$

$\Rightarrow \boxed{2000 = f_3 + f_4 + f_5}$ — (III)

③ Median 3350 lies in the 3000-4000

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

$3350 = 3000 + \frac{1000}{f_4}\left(\frac{230}{2} - 20 + f_3\right)$

$350 = \frac{1000}{f_4}(115 - 20 + f_3)$

$f_4 = \frac{1000}{350}(95 + f_3)$

$f_4 = 2.85(95 + f_3)$

$f_4 = 271.42 + 2.85 f_3$

$f_4 + 2.85 f_3 = 271.42$ — (I)

④ mode

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

$= 3000 + 1000\left(\frac{f_4 - f_3}{2f_4 - f_3 - f_5}\right)$

$3400 = 3000 + 1000\left(\frac{f_4 - f_3}{2f_4 - f_3 - f_5}\right)$

$\frac{2}{5} = \frac{f_4 - f_3}{2f_4 - f_3 - f_5}$

$5f_4 - 5f_3 = 4f_4 - 2f_3 - 2f_5$

$f_4 - 3f_3 = -2f_5$ — (II)

From (I) & (IV):

$f_4 + 2.85 f_3 = 271.42$

$-4f_3 + f_5 = 2000$

$f_4 - 1.85 f_3 - f_5 = -1728.58$ — (V)

From (II) & (III):

$f_4 - 1.85 f_3 - f_5 = -1728.58$ — (V)

From (II) & (III):

$2000 - f_3 - f_5 - 3f_3 = -2f_5$

$-4f_3 + f_5 = 2000$ — (IV)

From (II) & (V):

$f_4 - 3f_3 + 2f_5 = 0$

$f_4 - 1.85 f_3 - f_5 = -1728.58$

$-1.15 f_3 + 3f_5 = 1728.58$ — (VI)

From (I) & (II):

$f_4 + 2.85 f_3 = 271.42$

$f_4 - 3f_3 = -2f_5$

$5.85 f_3 = 271.42 + 2f_5$ — (IV)

From (III) & (IV):

$-12 f_3 + 2f_5 = 6000$

$1.15 f_3 + 3f_5 = 1728.58$

$10.85 f_3 = 4721.42$

7th February 2021

lecture-16 summary

- Measure of Dispersion
- Range
- Range's coefficient
- Interquartile Range

Measure of Dispersion

- Range
- QD
- SD
- Mean Deviation

Range

Range = U − L

- upper limit (x_{max})
- lower limit (x_{min})

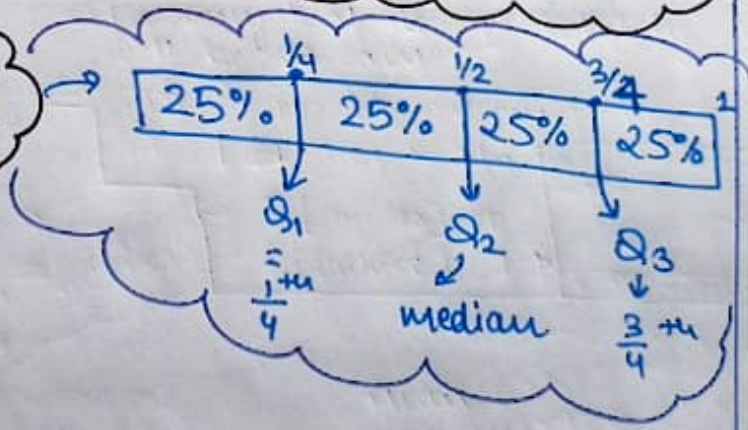
The yeilds of a cotton variety from 5 plots are 8, 9, 8, 10, 11. Find the Range!

Range = 11 − 8 = 3

INTERQUERTILE RANGE

→ Diff. b/w the upper & lower quartile values in the set of values

$$IQR = Q_3 - Q_1$$



$$COR = \frac{75-60}{75+60} \times 100$$

$$= \frac{15}{135} \times 100$$

$$= 11.11\%$$

Calculate Range from the following distribution

SIZE:	60-63	63-66	66-69	69-72
NUM.:	5	1	42	27

72 − 75

8

Range = 75 − 60 = 15

$$\text{coeff. of Range (COR)} = \frac{U-L}{U+L} \times 100$$

- Question on interquartile range
- Quartile for continuous range

INTERQUARTILE RANGE

Q 2, 8, 5, 10, 12

A ① Arrange in ascending order

2, 5, 8, 10, 12

② For discrete data set

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

$= \left(\frac{5+1}{4}\right)^{th}$ term

$= (1.5)^{th}$ term

$Q_1 = \frac{2+5}{2} = \frac{7}{2}$

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

$= \left(\frac{3}{4} \times 6\right)^{th}$ term

$= \frac{9}{2} = 4.5^{th}$ term

$= \frac{10+12}{2} = 11$

n = count the total no. of ~~dat~~ data given

Quartile for continuous case

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

$Q_1 = l + \frac{h}{f}\left(\frac{N}{4} - cf\right)$

$Q_3 = l + \frac{h}{f}\left(\frac{3}{4}N - cf\right)$

Q Find,
(i) Q_1 (ii) Q_2 (iii) Q_3
(iv) Range (v) Interqu. Range
of the following nos.

12, 5, 22, 30, 7, 86, 15, 14, 42, 53, 25, 65

A

5, 7, 12, 14, 15, 22, 25, 30, 36, 42, 53, 65

n = 12

① $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ term $= \frac{13}{4}^{th}$

$= (3.25)^{th}$

$= \frac{12+14}{2}$

$= 13$

चाहे 0.25 हो या 0.50 हो या 0.75 हो। हम हमेशा उन दो no. का Avg. लेंगे।

② $Q_2 = \left(\frac{n+1}{2}\right)^{th} = \frac{13}{2} = 6.5^{th}$

$= \frac{22+25}{2}$

$= 23.5$

③ $Q_3 = \frac{3}{4}(n+1) = \frac{13}{4} \times 3 = 9.75^{th}$

$= \frac{86+92}{2}$

$= \frac{68}{2}$

$= 39$

④ $IQR = Q_3 - Q_1$

$= 39 - 13$

$IQR = 26$

⑤ Range $= 65 - 5 = 60$

12th February 2024

- Variance in discrete case
- Quartile Deviation
- Normal Distribution
- Binomial Distribution

Lecture - 18 Summary

- Variance & Mean 1 Marker Formulae

$$\text{Var}(x) = \frac{1}{n}\sum f_i (x_i - \bar{x})^2 \quad \text{[For the discrete case]}$$

$$\text{Quartile deviation} = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

Question

The no. of vehicals sold by a major Maruti showroom in a day was recorded for 10 working days. The data is given as:

DAY	FREQ.
1	20
2	15
3	18
4	5
5	10
6	17
7	21
8	19
9	25
10	28

→ arrange as per freq. in ascending then find quartile range & coeff.

NORMAL DISTRIBUTION

Total No. of possible = 2^{n} (n → no. of trial)

If no. of trials is exceedingly large than the normal trials possibility. then we will use Binomial Distribution

BINOMIAL DISTRIBUTION

when trials are independent & Identical & distributive.

→ In every trial the probability will be same.

→ इसमें दो possible outcomes होते ही होते हैं। 'Yes' & 'No'

→ P [Success] = p

→ P [failure] = 1 - p

Prob. Mass Func.

→ $P(x=x) = {}^{n}C_x p^x q^{n-x}$

x = P(success)

Var $\left(\frac{x}{2} + c\right) = ?$

& Var $(x) = a$

$\rightarrow \frac{1}{2^2}(\text{Var } x) + \text{Var}(c)$

$\rightarrow \frac{1}{4}a + 0$

Mean $(x+5) = ?$

Mean $(x) = 4$

Var $(x+c) = a$?

If Var $(x) = a$ & x

$\rightarrow$ Var (x) + Var (c) = x

Mean $(2x+c)$ & Mean $(x) = a$

$\rightarrow$ Mean $(2x)$ + Mean (c)

$\rightarrow 2a + c$

If mean $(x) = 4$

then mean $(x+5) = 4+5 = 9$

Mean $(x) + 5$

Var $(2x+c) = ?$ & Var $(x) = a$

$\rightarrow$ Var $(2x) + \text{Var}(c)$

$\rightarrow a^2$

13th February 2024

Lecture -19 Summary

Probability Mass function

→ $P(x=x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$

- p → Prob. of Success
- q → Prob. of failure

Mean(x) = np

var(x) = npq

mean(x) > var(x)
→ only in binomial dist.

Ques

Six dice are thrown 725 times. How many times do you want at least three dice to show 5 or 6.

Ans

$P[x \geq 3] \times 725$ $\quad n = 6$ dice

or

$(1 - P[x<3]) \times 725$

$1 - (P(x=0) + P(x=1) + P(x=2)) \times 725$

NORMAL DISTRIBUTION

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \left(e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right)$$;

$-\infty < x < \infty$
$\mu > 0$
$\sigma^2 > 0$

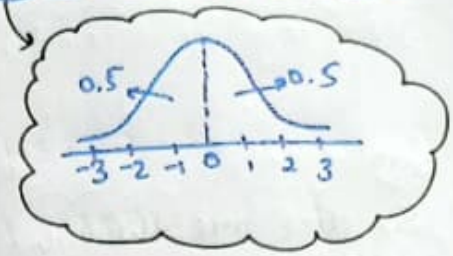
(Prob. Distribution Funct.)

Mean $(x) = \mu = E(x)$

Var $(x) = \sigma^2$

$X \sim N(\mu, \sigma^2)$

Also known as Gaussian Distribution



condition

- Mean, Median, Mode are equal
- The curve is symmetric about mean.
- Total area under the curves is 1
- ~~Left~~ Left hand side & Right hand side of the graph about the mean are same.

Z-Score

$$Z = \frac{X - \mu}{\sigma}$$

Prob. Density ~~Distrib~~ function.

$\mu = 0$

$\sigma = 1$

Question

If $z = \frac{x-\mu}{\sigma}$ then $f(z) = ?$

Answer

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}$$

- Questions on Normalisation

x is normally distributed & mean of x is 12 & Std. deviation is 4. & find out the prob. of the following

a) $(x \geq 20)$

b) $x \leq 20$

c) $0 \leq x \leq 12$

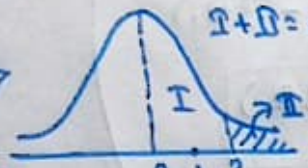
$Z = \frac{x-\mu}{\sigma}$

$X \sim N(12, 16)$

$P[x \geq 20] = P\left[\frac{x-\mu}{\sigma} \geq \frac{20-\mu}{\sigma}\right]$

$\downarrow$

$P[z \geq 2] \leftarrow P\left[z \geq \frac{20-12}{4}\right]$



$\rightarrow 0.5 - P[0 \leq z \leq 2]$ (given)

$\rightarrow 0.5 - 0.4772$

$\rightarrow 0.0228$

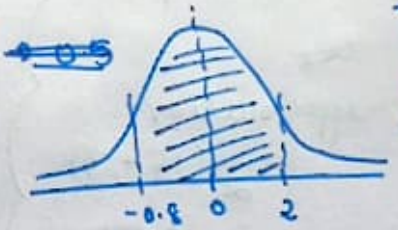
Range of the graph = -3 to +3

Q Mean : 30
SD = 5
$x \sim N(30, 25)$
Find the prob.
(i) $26 \leq x \leq 40$
(ii) $x \geq 45$
(iii) $|x-30| > 5$

$P[26 \leq x \leq 40]$

$P\left[\frac{26-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{40-\mu}{\sigma}\right]$

$P\left[\frac{-4}{5} \leq Z \leq 2\right]$



$P[-0.8 \leq x \leq 0] + P[0 \leq z \leq 2]$

$P[0 \leq x \leq 0.8] + P[0 \leq z \leq 2]$

$0.2881 + 0.4972$

$\simeq 0.78853$

coeffiecient of variation = $\frac{\sigma}{\bar{x}} \times 100$

σ = Std. Deviation

$\bar{x}$ = Mean

To find the outliers :

↳ lower outlier = $Q_1 - 1.5\,IQR = Q_1 - 1.5(Q_3 - Q_1)$

↳ upper outlier = $Q_3 + 1.5(IQR)$

$= Q_3 + 1.5(Q_3 - Q_1)$

$= Q_3(2.5) - (1.5Q_1)$

$= \frac{1}{10}(25Q_3 - 15Q_1)$

$= \frac{5}{10}(5Q_3 - 3Q_1)$

upper outlier = $\frac{1}{2}(5Q_3 - 3Q_1)$

Ques: 1, 3, 5, 8

Ans: $Q_3 = \frac{3}{4} \times 4 = 3$

$= 5$

$Q_1 = \frac{1}{4} \times 4 = 1$

$= 1$

upper outlier

$= \frac{1}{2}(15 - 3)$

$= \frac{12}{2} = 6$

upper outlier

$= 5 + 1.5(4)$

$= 5 + 6$

$= 11$

$= Q_1 - 1.5Q_3 + Q_1(1.5)$

$= +Q_1(0.5) - 1.5Q_3$

$= \frac{5}{10}(Q_1 - 3Q_3)$

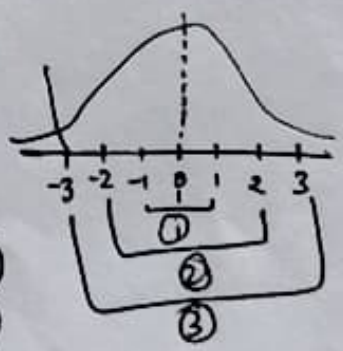
lower outlier = $\frac{1}{2}(Q_1 - 3Q_3)$

Area property of Normal dist

① $[\mu - \sigma < x < \mu + \sigma] = 0.6826$

② $[\mu - 2\sigma < x < \mu + 2\sigma] = 0.9544$

③ $[\mu - 3\sigma < x < \mu + 3\sigma] = 0.9973$



20th February 2024

- Imp. Qⁿ/Topics

→ Types & Data & variables with real life eg. (long answer type)

→ Graphical representation of data (1-2 Qⁿ)

VV Imp → Central Tendancy

→ Measures of Deviation (1-2 Qⁿ) → 100%.

→ Outliers (1 Qⁿ short Ans. type)

→ Outliers using interquartile range

→ Normal Distribution (using Z-score) for find probability.

$Var(x+y) = Var(x) + Var(y) - 2Cov(x,y)$

Note: If x & y are independent (diff) then covar. of (x & y) are zero.

COV

$\frac{\sigma}{\bar{x}} \times 100 = COV$

An analysis of monthly salary pay to the workers of 2 firms A & B belonging to the same industry gives the following results.

	FIRM A	FIRM B
No. of worker:	500	600
Avg. Salary:	186	175
variance of distribution:	81	100
Std. Deviation:	9	10

(i) Which firm A or B has a larger salary bill?

Firm A: 186 × 500 = 93000

Firm B: 600 × 175 = 105000

(ii) In which firm A or B is there greater variability in individual salary?

Firm A

$COV = \frac{9}{180} \times 100$

$= 4.83$

Firm B

$COV = \frac{10}{175} \times 100$

$= 5.71$