Statistics

31.08.24

- Raw data
- Arrays

Averages:

Arithmetic mean (A.M.) : $f_1, f_2, f_3 \dots$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{\Sigma f X}{\Sigma f}$$ long method

$$d = X - A$$

$$\overline{X} = A + \frac{\Sigma f d}{\Sigma f} \longrightarrow$$ short method

Coding method or step derivation method:

$$\overline{X} = A + \frac{\Sigma f u}{\Sigma f} i \text{ , where } u = \frac{X - A}{i}$$

* Find the arithmetic mean (A.M.) for the following distribution by step derivation method.

class frequency	frequency	Mid value X	$u = \frac{X-A}{i}$	fu
0 – 10	7	5	−2	−14
10 – 20	8	15	−1	−8
20 – 30	20	25	0	0
30 – 40	10	35	1	10
40 – 50	5	45	2	10
	50			−2

i = size of the class interval = difference b/w class boundaries $= 10$

A = the assumed mean

$$\bar{X} = 25 + \frac{-2}{50} \times 10$$
$$= 25 - \frac{2}{5}$$
$$= 24.6$$

0 – 9
10 – 19

এমন থাকলে graph plot এ ঝামেলা হবে
class boundaries column add করা লাগবে।

class boundaries	class frequency	frequency	Mid value X	$u = \frac{X-A}{i}$	fu
-0.5 – 9.5	0 – 9	7	5	-2	-14
9.5 – 19.5	10 – 19	8	15	-1	-8
19.5 – 29.5	20 – 29	20	25	0	0
29.5 – 39.5	30 – 39	10	35	1	10
39.5 – 49.5	40 – 49	5	45	2	10
		50			-2

Median:

5 7 8 12 15

5 7 $\boxed{8 \quad 12}$ 15 18

$$\frac{8+12}{2}$$

For grouped data:

$$\text{Median} = L + \left(\frac{\frac{N}{2} - c}{f}\right) i$$

where, L is the lower class boundary of the median class

N is the total frequency

c is the sum of the frequencies of all classes lower than the median class

i is the size of the median class

Mode:

series of data র মধ্যে যেটার frequency বেশি

2, 2, 5, 6, 6, 6, 9, 12, 15 → 6

2, 5, 7, 12 → mode নাই

(at least দুইবার থাকা লাগবে)

2, 2, 5, 5, 7, 7, 12, 12 → 2, 5, 7, 12

(more than 1 বার আছে)

07/09/24

$$\text{Median} = L_1 + \left(\frac{N/2 - F}{f}\right) i$$

where L_1 is the lower class boundary of the median class
N is the total frequency of the data
F is the sum of the frequencies of all classes lower than the median class.
f is the frequency of the median class.
and i is the width of the median class.

Find the median value from the following data.

median group

class interval :	0–5	5–10	$\boxed{10–15}$	15–20	20–25	25–30	30–35	35–40	40–45
frequency :	29	195	$\boxed{241}$	117	52	10	6	3	2

Total = 655

$$\frac{N}{2} = \frac{655}{2} = 327.5$$

10–15 এ আছে।

$L_1 = 10$

$F = 19 + 195 = 224$

$f = 241$

$i = 5$

$$\text{Median} = L_1 + \left(\frac{N/2 - F}{f}\right) i$$

$$= 10 + \left(\frac{655/2 - 224}{241}\right) \times 5$$

$$= 12.147$$

10–15 এর মধ্যে
($\therefore$ correct)

For grouped data :

$$\text{Mode} = L_1 + \left(\frac{f - f_1}{2f - f_1 - f_2}\right) i$$

where, L_1 is the lower class boundary of the modal class

f is the frequency of the modal class

f_1 is the frequency before the modal class frequency.

f_2 is the frequency after the modal class frequency.

Find the mode from the following data:

modal group

class interval :	0-6	6-12	13-18	$\boxed{18-24}$	24-30	30-36	36-42
frequency	6	11	25	$\boxed{35}$	18	12	6

$L_1 = 18$, $f = 35$, $f_1 = 25$, $f_2 = 18$, $i = 6$

$$\text{Mode} = L_1 + \left(\frac{f - f_1}{2f - f_1 - f_2}\right) i$$

$$= 18 + \frac{35 - 25}{2 \times 35 - 25 - 18} \times 6$$

$$= 20.222$$

২টার frequency 35 থাকলে how? — next class

The mean of 200 items was 50. Later on it was discovered that 2 items were misread as 92 and 8 instead of 192 and 88. Find the correct mean.

Soln:

$$\text{mean} = 50$$

$$\text{sum} = 200 \times 50$$

$$\text{correct sum} = 200 \times 50 - 92 - 8 + 192 + 88$$

$$= 10180$$

$$\therefore \text{correct mean} = \frac{10180}{200} = 50.9$$

group data র মধ্যে যদি কোনোটার frequency missing থাকে তাহলে কীভাবে?

From the following data find the missing frequency.

class interval :	40–43	43–46	46–49	49–52	52–55
Frequency :	31	58	60	?	27

It is given that the mean of the frequency is 47.2.

Soln:

class interval	f	mid value	u	fu	d	fd
40 – 43	31	41.5	−2	−62	−6	−186
43 – 46	58	44.5	−1	−58	−3	−174
46 – 49	60	47.5 (A)	0	0	0	0
49 – 52	x	50.5	1	x	3	$3x$
52 – 55	27	53.5	2	54	6	162
	$176 + x$			$-66 + x$		$-198 + 3x$

$$\overline{X} = A + \frac{\Sigma fu}{\Sigma f} i$$

$$47{\cdot}2 = 47{\cdot}5 + \frac{-66+x}{176+x} \times 3$$

$$\rightarrow x = 44$$

<u>or</u>

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

$$\rightarrow 47{\cdot}2 = 47{\cdot}5 + \frac{-198+3x}{176+x}$$

$$\therefore x = 44$$

mean of the frequency $47{\cdot}2 \rightarrow$

$$\frac{176+x}{5} = 47{\cdot}2$$

$$\rightarrow x = 60$$

11.09.24

Standard deviation is defined as the square root of the mean square of the deviation from the A.M.

$$S.D = \sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} \quad , \quad \sigma^2 = \text{variance}$$

Standard deviation zero হলে কি বোঝায়?

Calculate the S.D. for the following data:

class interval	f	x	$d = x - A$ $(A=6)$	fd	fd^2
0−4	4	2	−4	−16	64
4−8	8	6 (A)	0	0	0
8−12	2	10	4	8	32
12−16	1	14	8	8	64
	$\Sigma f = 15$			$\Sigma fd = 0$	$\Sigma fd^2 = 160$

$$\boxed{S.D = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}} = 3.266$$

যে formula দিয়েই করি same আসবে।

Following are the marks obtained by two students X and Y in 10 tests of 100 marks each:

marks obtained by X :	44	80	76	48	52	72	68	56	60	54
marks obtained by Y :	48	75	54	60	63	69	72	51	57	66

If the consistency of performance is the criterion for awarding a prize, who should get the prize?

$$\bar{x} = \frac{610}{10} = 61 \qquad \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = 11.7 \qquad \frac{\sigma}{\bar{x}} \times 100 = 19.18$$

$$\bar{y} = \frac{615}{10} = 61.5 \qquad \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = 8.63 \qquad \frac{\sigma}{\bar{x}} \times 100 = 14.03$$

(Winner)

coefficient of variation (C.V) $= \frac{\sigma}{\bar{x}} \times 100$

যার C.V. কম $\longrightarrow$ তার performance ভালো।

Ref:

R.S. Pillai
V. Bagavathi } statistics

Moments:

The rth moment of a variable x about the mean $\bar{x}$ is usually denoted by μ_r and is given by

$$\mu_r = \frac{1}{N} \Sigma f_i (x-\bar{x})^r, \quad \text{where } \Sigma f_i = N$$

$r = 1 \rightarrow$ first moment about the mean

$\rightarrow \mu_1$ will be always zero

x	$x - \bar{x}$
2	-1
3	0
4	1

$\bar{x} = 3$

The rth moment of a variable x about any point a is defined by

$$\mu_r' = \frac{1}{N} \sum f_i (x_i - a)^r$$

Relation between moments about mean and moment about any point a

$$\mu_1 = 0 = \mu_1' - \mu_1'$$

$$\mu_2 = \mu_2' - \mu_1'$$

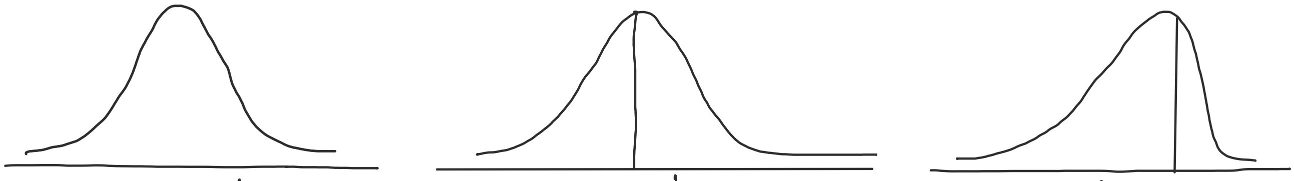
$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1' - 3\mu_1'^4$$

14.09.24

Skewness is the lack of symmetry. The measures of asymmetry are usually called measures of skewness

$$\text{coefficients of skewness} = \frac{\text{Mean} - \text{Mode}}{\text{s.d.}}$$



Measures of skewness : $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

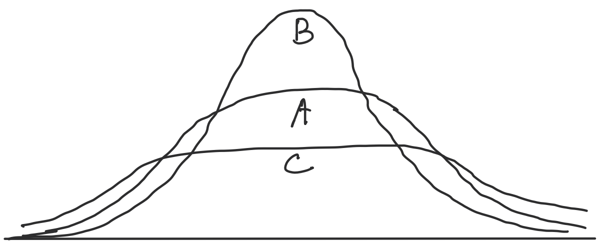
$\mu_4 \rightarrow$ 4th moment about the mean

Kurtosis: It measures the degree of peakedness of a distribution and is given by

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3, \text{ the curve is normal or mesokurtic distribution}$$

$\mu_2 \rightarrow$ 2nd moment about the mean

If $\beta_2 > 3$, the distribution is leptokurtic
and if $\beta_2 < 3$, the distribution is platykurtic distribution.



A — mesukurtic

B — lepto kurtic

C — platy kurtic

⊛ In a certain distribution the first four moments about the value 5 are 2, 20, 40 and 50. Calculate β_1 and β_2 and comment on the shape or nature of the distribution.

about any value a (μ')

Solution: Here, $\mu_1' = 2$, $\mu_2' = 20$, $\mu_3' = 40$ and $\mu_4' = 50$

$$\mu_1 = \mu_1' - \mu_1' = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 16$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3 = -64$$

$$\mu_4 = 162$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 1 \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} = 0.6328$$

→ (+ve) distribution ($\beta_1 > 0$) → platykurtic ($\beta_2 < 3$)

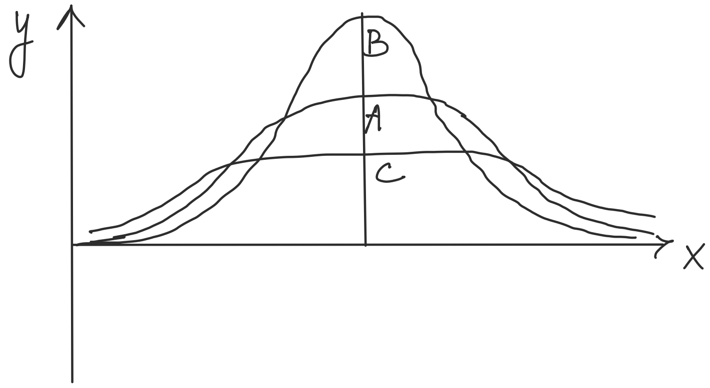
18.09.24

Kurtosis: The degree of Kurtosis of a distribution is measured relative to the peakedness or a normal

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3,$$ the curve is normal or mesokurtic distribution

$\mu_2 \to$ 2nd moment about the mean

If $\beta_2 > 3$, the distribution is leptokurtic
and if $\beta_2 < 3$, the distribution is platykurtic distribution.



A – mesukurtic
B – lepto kurtic
C – platy kurtic

* Calculate the first four moments from the following data and find out β_1 and β_2 and also comment on your result.

x	f	$f(x)$	$A=4$ $d = x-A$	$f(x-4)$	$f(x-4)^2$	$f(x-4)^3$	$f(x-4)^4$
0	5						
1	10						
2	15						
3	20						
4	25						
5	20						
6	15						
7	10						
8	5						
	125 $= N = \Sigma F$	$\Sigma fx = 500$		$\Sigma fd = 0$	$\Sigma fd^2 = 500$	$\Sigma fd^3 = 0$	$\Sigma fd^4 = 4700$

$\mu_1 = \frac{\Sigma fd}{N} = 0$ $\quad \mu_2 = \frac{\Sigma fd^2}{N} = \frac{500}{125} = 4$ $\quad \mu_3 = \frac{\Sigma fd^3}{N} = 0$ $\quad \mu_4 = \frac{\Sigma fd^4}{N} = 37.6$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0 \quad (\text{symmetric distribution})$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 2.37 < 3 \quad (\text{platykurtic})$$

Quartiles, Deciles and Percentiles:

(3টি) (9টি) (99টি)

First quartile, $Q_1 = L_1 + \dfrac{N/4 - c.f.}{f_{Q_1}} \times c$

f_{Q_1} → frequency of the first quartile class.

$$Q_2 = L_2 + \frac{2N/4 - c.f.}{f_{Q_2}} \times c$$

L_2 → lower class boundary of the second quartile class

$$Q_3 = L_3 + \frac{3N/4 - c.f.}{f_{Q_3}} \times c$$

2nd quartile = 5th decile = 50th percentile

6th decile, $D_6 = L_6 + \dfrac{6N/10 - c.f.}{f_{D_6}} \times c$

L_6 → lower class boundary of the 6th decile class.

f_{D_6} → frequency of the 6th decile class nth percentile এ $\frac{Nn}{100}$

$$65\text{th percentile} = L_{65} + \frac{65N/100 - c.f.}{f_{65}} \times c$$

$c \rightarrow$ interval

$c.f. \rightarrow$ ঐ class এর আগের class পর্যন্ত frequency র sum.

21.09.24

Probability: If an event A can happen in m ways, and failed to happen in n ways, all this ways being equally likely to occur then the probability of happening A is

$$= \frac{\text{number of favourable cases}}{\text{Total number of mutually exclusive and equally likely cases}} = \frac{m}{m+n} = p$$

and that the probability of not happening $= \frac{n}{m+n} = q$

$$p + q = 1$$

mutually exclusive — H আসলে T আসবে না and vice versa.

Dice: (more than 1 Die)

Die — একটা ছক্কা

Find the probability of throwing 9 with two dice.

↑ sum of the upper face of the dice.

sum = 9

(5, 4)
(4, 5)
(6, 3)
(3, 6)

$\therefore$ probability $= \frac{4}{36}$

$= \frac{1}{9}$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Expected Value: If $p_1, p_2, p_3 \ldots, p_n$ are probabilities of the events $x_1, x_2 \ldots, x_n$ respectively then expected value:

$$E(x) = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n = \sum_{i=1}^{n} p_i x_i$$

Conditional Probability: The probability of happening an event A, such that event B has already happened, is called the probability of happening A on the condition that B has already happened. It is usually denoted by $P(A|B)$.

If two events are mutually exclusive, then the probability of the occurrance of either A or B is the sum of the probabilities of A and B.

Thus $P(A \text{ or } B) = P(A) + P(B)$

Ex: A bag contains 3 white, 2 black and 5 red balls. What is the probability of getting a white or red ball at random in a single draw.

Solution: The probability of getting a white ball is $= \frac{3}{10}$

The probability of getting a red ball is $= \frac{5}{10}$

$\therefore$ The probability of getting a white or red ball is $= \frac{3}{10} + \frac{5}{10}$

$$= \frac{8}{10}$$

$$= \frac{4}{5}$$

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

When events are not mutually exclusive:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \cap B)$$

২ জনকে একটা problem দেওয়া হলো।

X এর solve করার probability $3/4$

Y " " " " $2/3$

problem টা solve হওয়ার probability = ?

Soln: $P(X \text{ or } Y) = P(X) + P(Y) - P(X \cap Y)$

$$= \frac{3}{4} + \frac{2}{3} - \frac{3}{4} \times \frac{2}{3}$$

$$= \frac{11}{12}$$