Statistics

By
Ankush Garg(B. Tech, IIT Jodhpur)

MEASURES OF CENTRAL TENDENCY

An average value or a central value of a distribution is the value of variable which is representative of the entire distribution, this representative value are called the measures of central tendency.

→ Mean
→ Median
→ Mode

1 ARITHMETIC MEAN :

(i) **For ungrouped dist.** : If $x_1, x_2, \ldots\ldots x_n$ are n values of variate x_i then their A.M. $\bar{x}$ is defined as

Average

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

$$\Rightarrow \qquad \Sigma x_i = n\,\bar{x}$$

$\dfrac{\quad}{50}$

x_i	f_i
50	5
62	11
75	12

(ii) **For ungrouped and grouped freq. dist.** : If $x_1, x_2, \ldots\ldots x_n$ are values of variate with corresponding frequencies $f_1, f_2, \ldots f_n$ then their A.M. is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{N}, \quad \text{where } N = \sum_{i=1}^{n} f_i$$

3). Find Mean of the distribution

x_i	f_i
40	2
20	7
30	8
40	3

$$\bar{x} = \frac{(40\times 2) + (20\times 7) + (30\times 8) + (40\times 3)}{20}$$

$$= \frac{8 + 14 + 24 + 12}{2} = \frac{58}{2} = 29$$

(iii) By short method : If the value of x_i are large, then the calculation of A.M. by using previous formula is quite tedious and time consuming. In such case we take deviation of variate from an arbitrary point a.

Let $$d_i = x_i - a$$

$$\therefore \quad \bar{x} = a + \frac{\Sigma f_i d_i}{N},$$ where a is assumed mean

$d_i = x_i - 200$

=

$\Rightarrow d_i$ $\frac{30}{10}$ $\frac{20}{10}$

10 50

(iv) By step deviation method : Sometime during the application of short method of finding the A.M. If each deviation d_i are divisible by a common number h(let)

Let $$u_i = \left(\frac{d_i}{h}\right) = \frac{x_i - a}{h}$$

$u_i =$ 3 2

1 5

$$\therefore \quad \bar{x} = a + \left(\frac{\Sigma f_i u_i}{N}\right) h$$

9) If for some $x \in R$, the frequency distribution of the marks obtained by 20 students in a test is

Marks	2	3	5	7
Frequency	$(x+1)^2$	$2x-5$	x^2-3x	x

Then, the mean of the marks is (2019 Main, 10 April I)

(a) 3.0 (b) 2.8 (c) 2.5 (d) 3.2

2	3	5	7
16	1	0	3

$$\bar{x} = \frac{32+3+21}{20} = \frac{56}{20} = 2.8$$

$$(x+1)^2 + (2x-5) + (x^2-3x) + x = 20$$

$$x^2 + 2x + 1 + 2x - 5 + x^2 - 3x + x = 20$$

$$2x^2 + 2x - 24 = 0$$

$$x^2 + x - 12 = 0$$

$$(x+4)(x-3) = 0$$

$$\boxed{x = 3}$$

If $\bar{x}$ is the mean of variate x_i then

A.M. of $(x_i + \lambda) = \bar{x} + \lambda$

A.M. of $(\lambda x_i) = \lambda \bar{x}$

A.M. of $(ax_i + b) = a\bar{x} + b$ (where λ, a, b are constant)

$x_1+\lambda, x_2+\lambda, x_3+\lambda$ - - - $x_n+\lambda$

$\bar{x}$?

$x_1, x_2,$ - - $x_n \longrightarrow \bar{x}$

ax_1+b, ax_2+b - - - $\longrightarrow a\bar{x}+b$

MEDIAN :

The median of a series is the value of middle term of the series when the values are written in ascending order. Therefore median, divided an arranged series into two equal parts.

Formulae of median :

(i) **For ungrouped distribution :** Let n be the number of variate in a series then

$$\text{Median} = \begin{cases} \left(\dfrac{n+1}{2}\right)^{th} \text{ term, (when n is odd)} \\ \text{Mean of} \left(\dfrac{n}{2}\right)^{th} \text{ and } \left(\dfrac{n}{2}+1\right)^{th} \text{ terms, (when n is even)} \end{cases}$$

$x_1 \ x_2 \ x_3 \ x_4 \ x_5$

$\left(\frac{5+1}{2}\right)^{th}$ 3^{rd}

$x_1, x_2 - - - - x_{10}$

$x_1 < x_2 < - - - - - x_{10}$

$x_1 \ x_2 \ x_3 \ x_4$

$\frac{x_2 + x_3}{2} = \text{Median}$

(ii) For ungrouped freq. dist. : First we prepare the cumulative frequency (c.f.) column and Find value of N then

$$\text{Median} = \begin{cases} \left(\dfrac{N+1}{2}\right)^{th} \text{ term, (when N is odd)} \\ \text{Mean of } \left(\dfrac{N}{2}\right)^{th} \text{ and } \left(\dfrac{N}{2}+1\right)^{th} \text{ terms, (when N is even)} \end{cases}$$

x_i	f_i	c.f.
x_1	f_1	f_1
x_2	f_2	f_1+f_2
x_3	f_3	$f_1+f_2+f_3$
		1

$\sum f_i = N$

$\sum f_i = N$ 101

$\left(\dfrac{101+1}{2}\right)^{th}$

51^{th}

(iii) **For grouped freq. dist :** Prepare c.f. column and find value of $\frac{N}{2}$ then find the class which contain value of c.f. is equal or just greater to N/2, this is median class

$$\therefore \boxed{\text{Median} = \ell + \frac{\left(\frac{N}{2} - F\right)}{f} \times h}$$

$\left(\frac{N}{2}\right)$

where ℓ – lower limit of median class

f – freq. of median class

F – c.f. of the class preceeding median class

h – Class interval of median class

x_i	f_i	c.f
10 – 20	10	
20 – 30	12	x
30 – 40	29	

$\sum f_i = N$

10

F

Median class

9) The mean and the median of the following ten numbers in increasing order 10, 22, 26, 29, 34, x, 42, 67, 70, y are 42 and 35 respectively, then $\frac{y}{x}$ is equal to

(2019 Main, 9 April II)

(a) $\frac{7}{3}$

(b) $\frac{7}{2}$

(c) $\frac{8}{3}$

(d) $\frac{9}{4}$

MODE :

In a frequency distribution the mode is the value of that variate which have the maximum frequency

Method for determining mode :

(i) **For ungrouped dist. :** The value of that variate which is repeated maximum number of times

(ii) **For ungrouped freq. dist. :** The value of that variate which have maximum frequency.

(iii) **For grouped freq. dist. :** First we find the class which have maximum frequency, this is model calss

$$\therefore \boxed{\text{Mode} = \ell + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times h}$$

where
- ℓ – lower limit of model class
- f_0 – freq. of the model class
- f_1 – freq. of the class preceeding model class
- f_2 – freq. of the class succeeding model class
- h – class interval of model class

$h = 10$

x_i	f_i	
10 – 20	10	
20 – 30	12	$\to f_1$
30 – 40	15	$\to$ Modal class f_0
40 – 50	7	$\to f_2$

RELATION BETWEEN MEAN, MEDIAN AND MODE

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

MEASURES OF DISPERSION :

8

The dispersion of a statistical distribution is the measure of deviation of its values about the their average (central) value.

(ii) Mean deviation (M.D.) : The mean deviation of a distribution is, the mean of absolute value of deviations of variate from their statistical average (Mean, Median, Mode).

If A is any statistical average of a distribution then mean deviation about A is defined as

$$\text{Mean deviation} = \frac{\sum_{i=1}^{n} |x_i - A|}{n} \qquad \text{(for ungrouped dist.)}$$

48 50 47

$$\text{Mean deviation} = \frac{\sum_{i=1}^{n} f_i |x_i - A|}{N} \qquad \text{(for freq. dist.)}$$

$$\boxed{N = \Sigma f_i}$$

$$\begin{array}{cccc} f_1 & f_2 & & \\ x_1, & x_2, & \ldots & x_n \end{array}$$

$$|x_1 - A|, |x_2 - A| \ldots |x_n - A|$$

$$\frac{f_1|x_1 - A| + f_2|x_2 - A| \ldots}{\Sigma f_i}$$

A → Mean, Median, Mode

9) If the sum of the deviations of 50 observations from 30 is 50, then the mean of these observations is

(2019 Main, 12 Jan I)

(a) 50 (b) 30 (c) 51 (d) 31

$x_1 - - - - x_{50}$

$(x_1 - 30) + (x_2 - 30) - - (x_{50} - 30) = 50$

$x_1 + x_2 - - x_{50} - (30 \times 50) = 50$

$x_1 + x_2 - - x_{50} = 50 + 30 \times 50$

$= 31 \times 50$

$\bar{x} = \frac{x_1 + x_2 - - x_{50}}{50}$

$= \frac{31 \times \cancel{50}}{\cancel{50}} = 31$

(iii) **Variance and standard deviation** : The variance of a distribution is, the mean of squares of deviation of variate from their mean. It is denoted by σ^2 or var(x).

The positive square root of the variance are called the standard deviation. It is denoted by σ or S.D.

Hence standard deviation = $+\sqrt{\text{variance}}$

σ^2, var(x)

S.D = $\sqrt{\text{variance}}$

$\boxed{S.D = \sigma}$

Formulae for variance :

(i) for ungrouped dist. :

$x_1 \quad x_2 \quad x_3 \quad x_4$

$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{4}$

$$\boxed{\sigma_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}}$$

$\sigma^2 = \frac{\Sigma (x_i - \bar{x})^2}{n}$

$$\boxed{\sigma_x^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2} = \frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2$$

$\sigma^2 = \frac{\Sigma x_i^2}{n} - (\bar{x})^2$

$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 - -}{4}$

$$\sigma_d^2 = \frac{\Sigma d_i^2}{n} - \left(\frac{\Sigma d_i}{n}\right)^2, \quad \text{where } d_i = x_i - a$$

$\frac{\Sigma\left[x_i^2 + (\bar{x})^2 - 2x_i\bar{x}\right]}{n}$

$\left(\frac{\Sigma x_i^2}{n}\right) + (\bar{x})^2 - 2\bar{x}\left(\frac{\Sigma x_i}{n}\right)$

Find the variance of first n natural numbers

$\underline{10}$, $\underline{15}$

$1, 2, \ldots n$.

$\bar{x} = \frac{(1+2+3 \ldots n)}{n}$

$= \frac{\cancel{n}(n+1)}{2\cancel{n}}$

$= \left(\frac{n+1}{2}\right)$

$\sigma^2 = \frac{\Sigma x_i^2}{n} - (\bar{x})^2$

$= \left(\frac{1^2+2^2 \ldots n^2}{n}\right) - \left(\frac{n+1}{2}\right)^2$

$= \frac{\cancel{n}(n+1)(2n+1)}{6\cancel{n}} - \frac{(n+1)^2}{4}$

$= (n+1)\left[\frac{2n+1}{6} - \frac{(n+1)}{4}\right]$

$\sigma^2 = (n+1)\left[\frac{4n+2-3n-3}{12}\right]$

$\sigma^2 = \frac{(n+1)(n-1)}{12}$

$\boxed{\sigma^2 = \frac{n^2-1}{12}}$

SAFALTA.COM

9) If the data $x_1, x_2, \ldots, x_{10}$ is such that the mean of first four of these is 11, the mean of the remaining six is 16 and the sum of squares of all of these is 2000, then the standard deviation of this data is **(2019 Main, 12 April I)**

(a) $2\sqrt{2}$ (b) 2 (c) 4 (d) $\sqrt{2}$

$$\frac{x_1+x_2+x_3+x_4}{4} = 11$$

$$x_1+x_2+\ldots x_4 = 44$$

$$x_5+x_6 \ldots x_{10} = 96$$

$$\sigma = \sqrt{\text{Variance}} \qquad \sigma = \sqrt{4}$$

$$\boxed{\sigma^2 = \frac{\Sigma(x_i^2)}{n} - (\bar{x})^2}$$

$$= \frac{2000}{10} - (\bar{x})^2$$

$$= 200 - (\bar{x})^2$$

$$= 200 - (14)^2$$

$$= 200 - 196$$

$$\sigma^2 = 4$$

$$\bar{x} = \frac{(x_1+x_2 \ldots x_{10})}{10}$$

$$= \frac{140}{10}$$

$$\bar{x} = 14$$

$x_1 \longrightarrow f_1$
$x_2 \longrightarrow f_2$

(ii) For freq. dist. :

$$\boxed{\sigma_x^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N}}$$

$$\sigma_x^2 = \frac{\Sigma f_i x_i^2}{N} - (\bar{x})^2 = \frac{\Sigma f_i x_i^2}{N} - \left(\frac{\Sigma f_i x_i}{N}\right)^2$$

$$\sigma_d^2 = \frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2$$

$$\sigma_u^2 = h^2\left[\frac{\Sigma f_i u_i^2}{N} - \left(\frac{\Sigma f_i u_i}{N}\right)^2\right] \quad \text{where } u_i = \frac{d_i}{h}$$

(iii) $\boxed{\text{Coefficient of S.D.} = \frac{\sigma}{\bar{x}}}$

$$\boxed{\text{Coefficient of variation} = \frac{\sigma}{\bar{x}} \times 100}$$

Note :- $\sigma^2 = \sigma_x^2 = \sigma_d^2 = h^2 \sigma_u^2$

$A^3 = O$

$(I - A)^{-1} = B$

$\Rightarrow I = B(I - A)$

$I = B - BA$

$A = B \cdot A - BA^2$

$A^2 = BA^2 - \cancel{BA^3}$

$BA^2 - A^2 = O$

$(B - I) \cdot A^2 = O$

$A^2 \neq O$

$B - I = O$

$B - I = 0$

$B = I$

9) The mean of five observations is 5 and their variance is 9.20. If three of the given five observations are 1, 3 and 8, then a ratio of other two observations is

(2019 Main, 10 Jan I)

(a) 4 : 9 (b) 6 : 7 (c) 10 : 3 (d) 5 : 8

$1, 3, 8, x, y$

$\frac{1+3+8+x+y}{5} = 5$

$12 + x + y = 25$

$\boxed{x + y = 13}$

$\sigma^2 = 9.2$

$9.2 = \frac{\sum x_i^2}{n} - (\bar{x})^2$

$9.2 = \frac{1+9+64+x^2+y^2}{5} - 25$

$34.2 \times 5 = 74 + x^2 + y^2$

$171 = 74 + x^2 + y^2$

$97 = x^2 + y^2$