Lecture no – 3 10/2/2024

10 February 2024 11:46

2] Weighted mean: A weighted mean is a kind of average. Instead of each data point contributing equally to the final mean, some data points contribute more "weight" than others.

To calculate an average that takes into account the importance of each value to the overall cost . Find out average cost of labor per hour for each of the products.

		Labor hrs per unit of output	
Grade of labor	Hourly wage	Product 1	Product 2
Unskilled	5	1	4
Semiskilled	7	2	3
Skilled	9	5	3
		8	10

7×1 + 7×2 + 7×5

Method 1: A simple arithmetic mean = (5+7+9) / 3= 7/hr

Using this, labor cost of 1 unit of product 1 to be = 7* (1+2+5) = 56

Product 2 = 7* (4+3+3) = 70

Both are incorrect, the answers must take into account that different amount of each grade of labor .

Method 2: Weighted Mean:

P1= avg cost of labor per hr = (5*1+7*2+9*5)/8 = 8

P2= avg cost of labor per hr = (5*4+7*3+9*3)/10 = 6.80

3] Geometric mean: Sometimes when we are dealing with quantities that change over a period of time , we need to know an average rate of change , such an average growth over a period of several years. In such cases, simple arithmetic mean is inappropriate, because it gives wrong answer.

Ex: Rs. 100 deposited in saving account.

Year Interest rate Growth factor Saving at the end of year

Year	Interest rate
1	7%
2	8
3	10
4	12
5	18

Solution:

$\frac{7}{100} = 0.07$ 1+0.07 = 1.07

Year	Interest rate	Growth factor	Saving at the end of year	
1	7%	1.07	107.00	= 1.07 × 100
2	8	1.08	115.56	= 107 × 1.08
3	10	1.10	127.116	= 115.56 × 1.10
4	12	1.12	142.37	= 127.116 × 1.12
5	18	1.18	168.00	

Mean of growth factor = 1.07+1.08+…..+1.18)/ 5 = 1.11, corresponds **to 11% rate.**

100* (1.11)*(1.11) (1.11)*(1.11)* (1.11) = **168.51,** correct **growth rate should be less than 1.11**.

$\Rightarrow \left[GM = \sqrt[n]{\text{Product of x values}} \right] = (1.07*1.08*..........*1.18)^{1/n} \Rightarrow 1/5$

= 1.1093 = 10.93%

↑ IR

Disadvantages of Mean:

It may be affected by extreme values → outlier

Tedious to compute

Cannot compute in case of open class

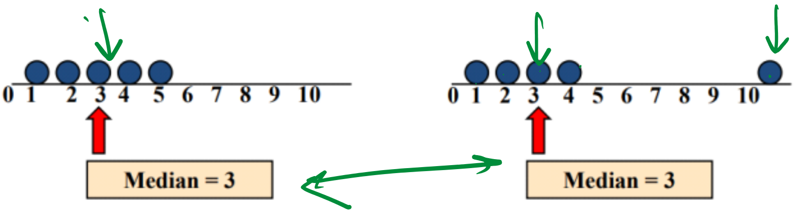
Cannot compute in case of categorical data

2] Measures of Central Tendency: The Median

1, 6, 7, 8, 10, 12, 9

- In an ordered array, the median is the "middle" number (50% above. 50% below)

- In an ordered array, the median is the "middle" number (50% above, 50% below)



- Not affected by extreme values

Measures of Central Tendency: Locating the Median

- The location of the median when the values are in numerical order (smallest to largest):

$$\text{Median position} = \frac{n+1}{2} \text{ position in the ordered data}$$

- If the number of values is odd, the median is the middle number
- If the number of values is **even**, the median is the **average of the two middle numbers**

ungrouped → 1 2 $\boxed{3\ 4}$ 5 6

$$\text{median} = \frac{3+4}{2} = 3.5$$

Note that is not the *value* of the median, only the *position* of the median in the ranked data.

B) Median for Grouped Data

Formula for Median is given by

$$\text{Median} = L + \frac{(n/2) - m}{f} \times c$$

no. of Classes mid points freq^n

Where

L = Lower limit of the median class

n = Total number of observations = $\sum f(x)$

m = Cumulative frequency preceding the median class

f = Frequency of the median class

c = Class interval of the median class

3 - 5 ← median → f

L

(C = 5 - 3 = 2) interval

Solution for the Example

Class	Frequency	Cumulative Frequency
0-1	1	1
1-2	4	5
2-3	8	13
3-4	7	20
4-5	3	23
5-6	2	25
Total	**25** → Σf	

median class: 2-3

Preceding class cumulative freq: 5

L = Lower limit of the median class
n = Total number of observations
m = Cumulative frequency **preceding** the median class
f = Frequency of the median class
c = Class interval of the median class

Substituting in the formula the relevant values,

$$\text{Median} = L + \frac{(n/2) - m}{f} \times c \quad \text{we have Median} = 2 + \frac{(25/2) - 5}{8} \times 1$$

= 2.9375 //

$= 2.9375$

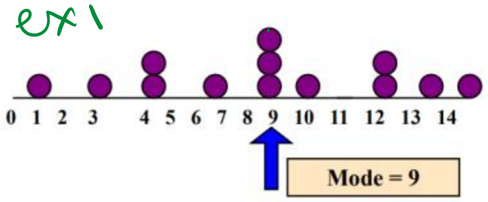
Advantages:
Not affected extreme values ✓
Can be computed in case of open class, if median is not in open class
Can be computed in case categorical variable

DisAd: Arraying of the data is time consuming.
To estimate population parameter, mean is easier.

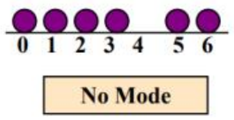
3) Measures of Central Tendency: The Mode

- Value that occurs most often
- Not affected by extreme values ← outliers
- Used for either **numerical or categorical** data
- There may be no mode
- There may be several modes

ex 1



ex 2



ungrouped

Mode for Grouped Data

$$\text{Mode} = L + \frac{d_1}{d_1 + d_2} \times c$$

Where L =Lower limit of the modal class

$d_1 = f_1 - f_0$ $\quad$ $d_2 = f_1 - f_2$

f_1 = Frequency of the **modal class**

f_0 = Frequency **preceding** the modal class

f_2 = Frequency **succeeding** the modal class. C = **Class Interval** of the modal class

Mode for Grouped Data Example

Example: Find the mode for the following continuous frequency distribution:

Class	0-1	1-2	2-3	3-4	4-5	5-6
Frequency	1	4	8	7	3	2

Solution for the Example

Class	Frequency
0-1	1
1-2	4
2-3	8
3-4	7
4-5	3
5-6	2
Total	**25**

modal class.

$$\text{Mode} = \boxed{L + \frac{d_1}{d_1 + d_2} \times c}$$

L = 2

$\boxed{d_1 = f_1 - f_0}$ = 8-4 = 4

$\boxed{d_2 = f_1 - f_2}$ = 8-7 = 1

C = 1 Hence Mode = $\boxed{2+\frac{4}{5}\times 1}$

= 2.8

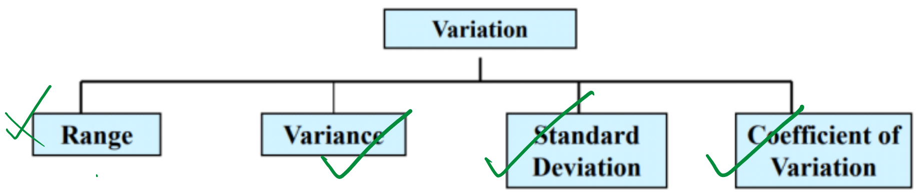
Measures of Central tendancy

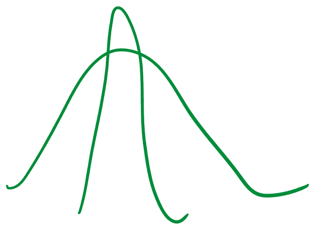
1 → Mean

2 → Median

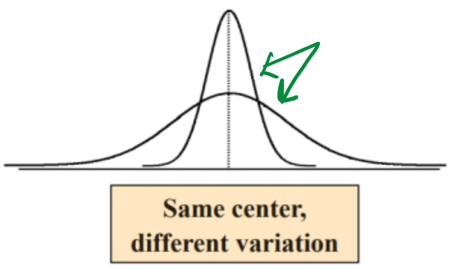
3 → mode

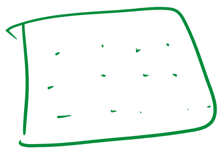
Measures of Variation





- Measures of variation give information on the **spread** or **variability** or **dispersion** of the data values.





1]

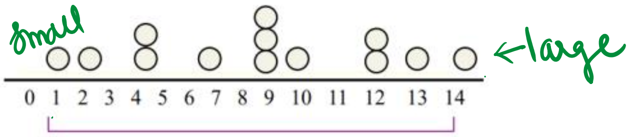
Measures of Variation: The Range

- Simplest measure of variation
- Difference between the largest and the smallest values:

$$\text{Range} = X_{\text{largest}} - X_{\text{smallest}}$$

$$\boxed{\text{Range} = X_{\text{largest}} - X_{\text{smallest}}}$$

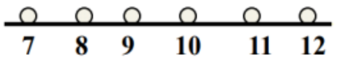
Example:



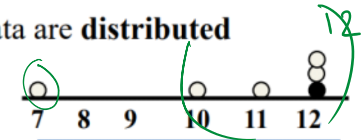
Range = 14 - 1 = 13

Measures of Variation:
Why The Range Can Be Misleading

- **Ignores** the way in which data are **distributed**



Range = 12 - 7 = 5



Range = 12 - 7 = 5

- Sensitive to outliers

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

Range = 5 - 1 = 4

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,**120** x

Range = 120 - 1 = 119

2]

Measures of Variation: The Variance

- Average (approximately) of squared deviations of values from the mean

– Sample variance: $$\boxed{S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

Where $\bar{X}$ = arithmetic mean

n = sample size

X_i = i^{th} value of the variable X

3]

Measures of Variation: The Standard Deviation

- **Most commonly** used measure of variation
- Shows variation about the **mean**
- Is the **square root of the variance**
- Has the same units as the original data

– Sample standard deviation: $S = \sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$

Measures of Variation: The Standard Deviation

Steps for Computing Standard Deviation

$\bar{x}$

1. Compute the **difference between each value** and the **mean**.
2. **Square** each difference.
3. **Add the squared** differences.
4. **Divide this total by n-1 to** get the sample variance.
5. Take the **square root of the sample** variance to get the sample standard deviation.

Example

Sample Data (X_i) : 10 12 14 15 17 18 18 24

ungrouped data.

$n = 8$ Mean $= \bar{X} = 16$

$$S = \sqrt{\frac{(10-\bar{X})^2 + (12-\bar{X})^2 + (14-\bar{X})^2 + \cdots + (24-\bar{X})^2}{n-1}}$$

$$= \sqrt{\frac{(10-16)^2 + (12-16)^2 + (14-16)^2 + \cdots + (24-16)^2}{8-1}}$$

$$= \sqrt{\frac{130}{7}} = 4.3095$$

Sample variance

Table :-

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

Observation (x) (1)	Mean ($\bar{x}$) (2)	$x - \bar{x}$ (1) − (2)	$(x - \bar{x})^2$ [(1) − (2)]^2	x^2 (1)^2
863	1,351	−488	238,144	744,769
903	1,351	−448	200,704	815,409
957	1,351	−394	155,236	915,849
1,041	1,351	−310	96,100	1,083,681
1,138	1,351	−213	45,369	1,295,044
1,204	1,351	−147	21,609	1,449,616
1,354	1,351	3	9	1,833,316

← table

903	1,351	−448	200,704	815,409
957	1,351	−394	155,236	915,849
1,041	1,351	−310	96,100	1,083,681
1,138	1,351	−213	45,369	1,295,044
1,204	1,351	−147	21,609	1,449,616
1,354	1,351	3	9	1,833,316
1,624	1,351	273	74,529	2,637,376
1,698	1,351	347	120,409	2,883,204
1,745	1,351	394	155,236	3,045,025
1,802	1,351	451	203,401	3,247,204
1,883	1,351	532	283,024	3,545,689
		$\Sigma(x-\bar{x})^2 \rightarrow$	1,593,770	23,496,182 $\leftarrow \Sigma x^2$

← table

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$= \frac{1{,}593{,}770}{11}$$

$= 144{,}888$ (or 144,888 [thousands of dollars]^2) ← Sample variance

$$s = \sqrt{s^2}$$

$$= \sqrt{144{,}888}$$

$= 380.64$ (that is, \$380,640) ← Sample standard deviation

Standard Deviation for Grouped Data

Standard Deviation (Sample) for Grouped Data

Frequency Distribution of Return on Investment of Mutual Funds

Return on Investment	Number of Mutual Funds
5-10	10
10-15	12
15-20	16
20-25	14
25-30	8
Total	**60**

Solution:

Solution for the Example

A	B	C	D	E	F	G	H
1	Return on	Investment		No of			
2			MidPoint	Funds			
3	Lower limit	Upper Limit	X	f	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
4	5	10	7.50	10	75	96.69	966.94
5	10	15	12.50	12	150	23.36	280.33
6	15	20	17.50	16	280	0.03	0.44
7	20	25	22.50	14	315	26.69	373.72
8	25	30	27.50	8	220	103.36	826.89
9				$n = \Sigma f =$ 60	Σ 1040		2448.33
10				Mean=	17.333		
11				Sample Variance =			41.50
12				Sample Standard Deviation =			6.44

$\frac{5+10}{2}$

$\frac{25+30}{2}$

$$\frac{\Sigma fx}{n} = \frac{1040}{60} = 17.33$$

$$s = \sqrt{\frac{\sum_{x=i}^{n}(x-\bar{x})}{n-1}}$$

$$s = \sqrt{\frac{\sum_{x=i}^{n}f(x-\bar{x})f}{n-1}}$$

From the spreadsheet of Microsoft Excel in the previous slide, it is easy to see

Mean = $\bar{X} = \frac{\sum fX}{n}$ =1040/60=17.333(cell F10),

Standard Deviation = s = $\sqrt{\frac{\sum f(X-\bar{X})^2}{n-1}}$ $\sqrt{\frac{2448.33}{59}}$

(Cell H12)

From the spreadsheet of Microsoft Excel in the previous slide, it is easy to see

$$S = \sqrt{\frac{\sum_{x=i}^{n} f(x-\bar{x}) f}{n-1}}$$

Mean = $\bar{X} = \frac{\sum fX}{n}$ =1040/60=17.333(cell F10),

Standard Deviation = S = $\sqrt{\frac{\sum f(X-\bar{X})^2}{n-1}}$ $\sqrt{\frac{2448.33}{59}}$
(Cell H12)

4]

Measures of Variation: Comparing Standard Deviations

The **coefficient of variation** (CV) is a measure of relative **variability**.

It is the ratio of the **standard deviation to the mean** (average).

Always in percentage (%)

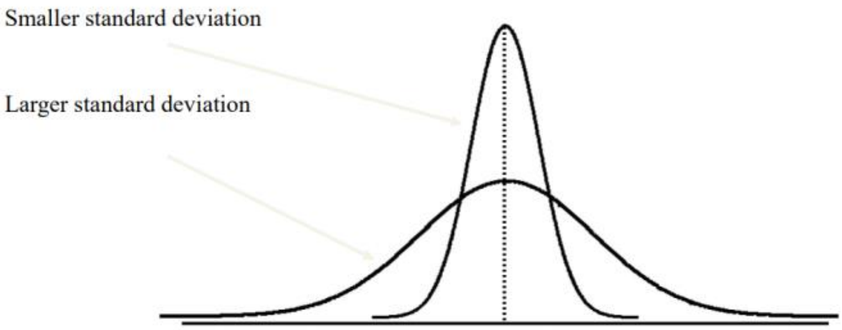
Shows **variation relative to mean**

Can be used to compare the variability of two or more sets of data measured in **different units**

$$CV = \left(\frac{S}{\bar{X}}\right) \cdot 100\% \leftarrow$$

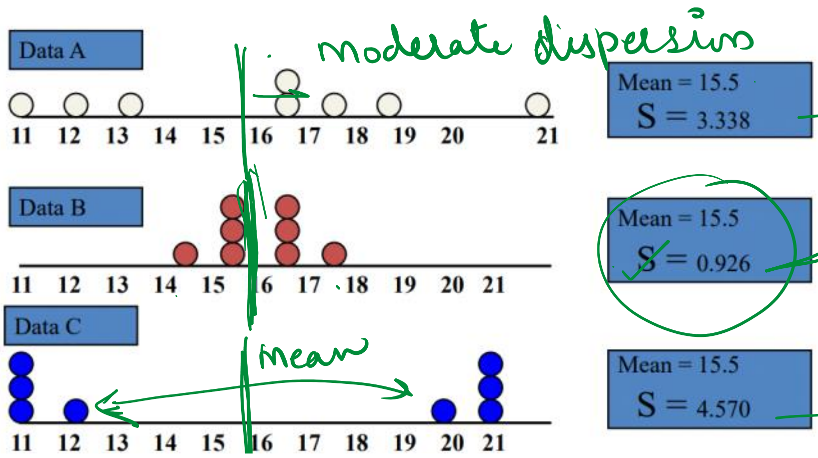
cv for sample

Compare standard deviation for two curves:

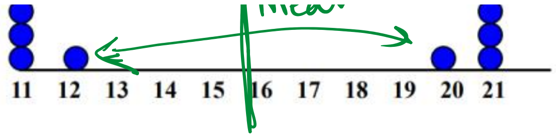


Measures of Variation: Comparing Standard Deviations

The **coefficient of variation** (CV) is a measure of relative **variability**. It is the ratio of the **standard deviation to the mean** (average).



mean



Mean = 15.5
S = 4.570

- Stock A:
  - Average price last year = \$50
  - Standard deviation = \$5

$$CV_A = \left(\frac{S}{\overline{X}}\right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

- Stock B:
  - Average price last year = \$100
  - Standard deviation = \$5

$$CV_B = \left(\frac{S}{\overline{X}}\right) \cdot 100\% = \frac{\$5}{\$100} \cdot 100\% = 5\%$$

Both stocks have the same standard deviation, but stock B is less variable relative to its price

① Range
② standard deviation
③ variance
④ C.V.

Sample statistics versus population parameters

S μ σ^2

Measure	Population Parameter	Sample Statistic
Mean	μ	$\overline{X}$
Variance	σ^2	S^2
Standard Deviation	σ	S

measures of central tendency & variation

Asg 3.