Numerical Descriptive Measures

The **central tendency** is the extent to which all the **data values group around a typical or central value**.

The **variation** is the amount of **dispersion**, or **scattering**, of values

The **shape** is the pattern of the distribution of values from the **lowest value to the highest** value.

Measures of Central Tendency: The Mean

- The arithmetic mean (often just called "mean") is the most common measure of central tendency
- For a sample of size n:

Pronounced x-bar

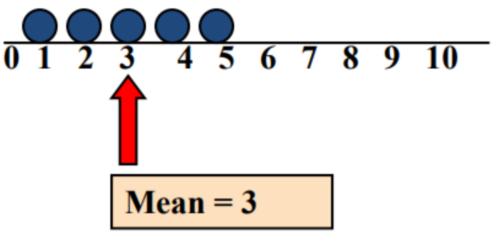
The i^{th} value

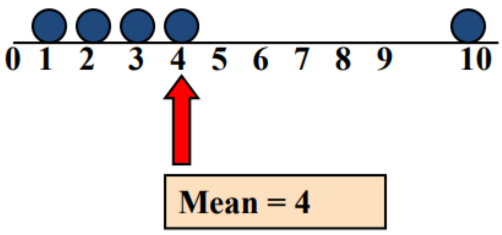
$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Sample size

Observed values

- The most common measure of central tendency
- Mean = sum of values divided by the number of values
- Affected by extreme values (outliers)





$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$

$$\frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$

Mean for Grouped Data

Formula for Mean is given by $\bar{X} = \frac{\sum f(X)}{n}$

Where

$\bar{X}$ = Mean

$\sum f(X)$ = Sum of cross products of frequency in each class with midpoint X of each class

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

Find the arithmetic mean 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

	A	B	C	D
1	Class	X (mid pt)	f	fX
2	0-1	0.5	1	0.5
3	1-2	1.5	4	6.0
4	2-3	2.5	8	20.0
5	3-4	3.5	7	24.5
6	4-5	4.5	3	13.5
7	5-6	5.5	2	11.0
8	Totals		25	75.5
9	Mean			3.02

Applying the formula $\bar{X} = \frac{\sum f(X)}{n}$ = 75.5/25=3.02

Mean using coding:

Class	f
0-7	2
8-15	6
16-23	3
24-31	5
32-39	2
40-47	2

$$\textbf{Mean} = x0 + w * \frac{(\text{Summation of } u*f))}{n}$$

w=numerical width of class interval
X0=value of midpoint assigned code 0

Mean using coding:

Class	mid	f	Code (u)	u*f
0-7	3.5	2	-2	-4
8-15	11.5	6	-1	-6
16-23	19.5	3	0	
24-31	??	5	1	??
32-39	??	2	2	4
40-47	43.5	2	3	6
		20		5

$$\textbf{Mean} = \mathbf{x0 + w * \frac{(Summation\ of\ u*f)}{n}}$$

= 19.5+8* (5) / (20) = 21.5

w=numerical width of class interval
X0=value of midpoint assigned code 0

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 f**or each of the product

		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

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
2 = 7* (4+3+3) = 70
Both are incorrect, the answers must take into account **that different amount of each grade of labor** .

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

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
1	7%	1.07	107.00
2	8	1.08	115.56
3	10	1.10	
4	12	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**.

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

$$= 1.1093 = 10.93\%$$

Disadvantages of Mean:

It may be affected by extreme values

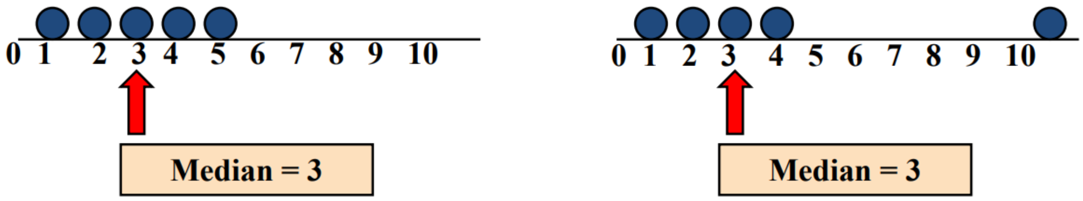
Tedious to compute

Cannot compute in case of open class

Cannot compute in case of categorical data

Measures of Central Tendency: The Median

- 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**

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

Ex:?????

Median for Grouped Data

Formula for Median is given by

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

Where

L =Lower limit of the median class

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

m = Cumulative frequency preceding the median class

f = Frequency of the median class

c = Class interval of the median class

Median for Grouped Data Example

Find the median 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	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**	

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,

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

= 2.9375

Find median

Class interval		f
0	49.99	78
50	99.99	123
100	149.99	187
150	199.99	82
200	249.99	51
250	299.99	47
300	349.99	13
350	399.99	9
400	449.99	6
450	499.99	4
		600

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

300^{th} = 126.17

301st=126.44

Find median

Class interval		f	Cum f
0	49.99	78	78
50	99.99	123	201
100	149.99	187	388
150	199.99	82	
200	249.99	51	
250	299.99	47	
300	349.99	13	
350	399.99	9	
400	449.99	6	
450	499.99	4	
		600	

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

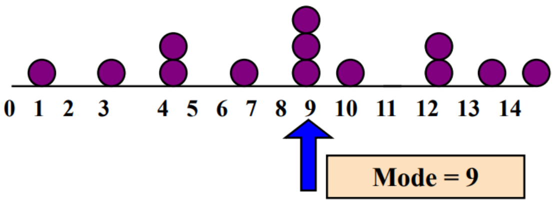
$$L + \frac{(n/2) - m}{f} \times c \qquad = 126.47$$

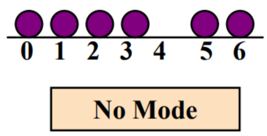
Advantages:
Not affected extreme values
Can be computed in case of open class, if median is not in open class
Can be computed in case categorical variable

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

Measures of Central Tendency: The Mode

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





Mode for Grouped Data

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

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**

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

$L = 2$

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

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

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

$= 2.8$