Components of numerical measures of data

When we speak of a data set, we refer to either a sample or to a population. If statistical inference is one's goal, s/he will wish ultimately to use to use sample numerical descriptive measures to make inferences about the corresponding measures for population. Although a large number of numerical methods are available to describe quantitative data sets. Most of these methods measures one of the two data characteristics:

Central tendency - This measures the extent to which all the values grouped around a typical or central value.

Variation or Dispersion–This measure the amount of dispersion or scattering of values away from a central value.

Measures of Central Tendency
Most sets of data show a distinct tendency to group around a central point. That is in a data set (population or sample) the values have a tendency to cluster around a certain point. This tendency of clustering the values around the center of the series is usually called central tendency. The numerical measure of this tendency of concentration is variously known as the measure of central tendency or measure of location or the measure of average.

Necessity of measuring the central tendency:
The necessities of measuring central tendency or average are as follows –

i. They give us an idea about the concentration of the values in the central part of the distribution.
ii. It is the value of the variable, which is typical of the whole set.
iii. It represents all relevant information contained in the data in as few numbers as possible.
iv. They give precise information, not information of a vogue general type.

Characteristics of a good measure of central tendency:
The following are the characteristics of an ideal measure of central tendency

i. It should be easy to understand.
ii. It should be easy to calculate.
iii. It should be based upon all observations.
iv. It should be rigidly defined.
v. It should be unduly affected by extreme values.
vi. It should be suitable for further algebraic treatment.
vii. It should be less affected by sampling fluctuation.

Different measure of central tendency:
The following are the different measure of central tendency

i. Arithmetic mean	ii. Median	iii. Mode
iv. Geometric mean	v. Harmonic mean	vi. Weighted mean

Arithmetic mean (AM):
Adding the values of the observations and then dividing the sum by the number of observations obtain the arithmetic mean of a series of observations.

Arithmetic mean (AM) for

- sample observation is denoted by $\bar{X}$
- Population mean is denoted by μ .

Suppose there are n values $x_1, x_2, \ldots, \ldots, \ldots, x_n$ for a variable X, then the AM denoted by $\bar{X}$ is defined as

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + \cdots + x_n}{n}; (i = 1,2, \ldots, n)$$

Example:
Banglatel is studying the number of minutes used by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.

90	77	94	89	119	112
91	110	92	100	113	83

What is the mean (arithmetic mean) number of minutes used?

Answer:
Average use of the rate plan

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + \cdots + x_n}{n} = \frac{90 + 77 + \cdots + 91 + \cdots + 113 + 83}{12} = 97.5$$

Thus the arithmetic mean number of minutes used last month by the sample of cell phone users is 97.5 minutes.

Again, for a ***group data*** as given in the following table

$$\begin{array}{rcccccc} Values: & x_1 & x_2 & \ldots & \ldots & x_k \\ Frequencies: & f_1 & f_2 & \ldots & \ldots & f_k \end{array}$$

Such that $f_1 + f_2 + \ldots + f_k = n$ then the AM is denoted by $\bar{X}$ is defined as

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \cdots + f_kx_k}{n}; (i = 1,2, \ldots, k)$$

Example & Exercise:
Calculate the mean for the following frequency distribution for n=100:

Class interval	Frequency
0-10	10
10-20	20
20-30	40
30-40	20
40-50	10

Answer:
Calculation:

Class interval	Frequency (f_i)	Mid values (x_i)	$(f_i) * (x_i)$
0-10	10	5	50
10-20	20	15	
20-30	40		
30-40	20		
40-50	10	45	450
Total	$\sum_{i=1}^{k=5} f_i =$		$\sum_{i=1}^{k=5} f_i * x_{i=} =$

Arithmetic mean

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \cdots + f_kx_k}{n}$$

$$= \frac{50 + \cdots + 450}{100} = ??$$

Test yourself

1. The following data represent the distribution of the age of employees within two different divisions of publishing company. Determine which company have relatively aged group of employees.

Age of employees	Number of employees of division	
	X	Y
20 – 30	6	13
30 – 40	19	30
40 – 50	9	24
50 – 60	10	0
60 – 70	2	4

When to use Arithmetic Mean:
In the following cases arithmetic mean should not be used:

1. In highly - skewed distributions.
2. In distributions with open
3. When the distribution is unevenly spread. Concentration being small or large at irregular points.
4. When an average rate of growth or change over a period of time is required.
5. When the observation are from geometric progression.
6. When averaging rates (that is speed, fluctuations in the prices of articles, etc.)
7. When there are very large and very small values of observations.

Hints: Fill the shaded area:

Age of employees	Number of employees of division		Mid value (m)	x*m	y*m
	X	Y			
20 – 30	6	13	25	150	325
30 – 40	19	30	35		
40 – 50	9	24	45		
50 – 60	10	0	55		
60 – 70	2	4	65	130	260
Total	46			1900	2715

AM of X division=	41.30435	**Obtained by (1900/46)**
AM of Y Division=		Calculate your self- ***result would be 38.23***

For any queries related to this presentation please contact
IFTEKHAR Mohammad Shafiqul Kalam
Assistant Professor
Department of Mathematics and Natural Sciences
Email: imskalam@bracu.ac.bd