CHAPTER TWO

2. Summarizing of data

2.1 Measure of central tendency

Objectives:

- ✓ To comprehend the data easily
- ✓ To facilitate comparison
- ✓ To make further statistical analysis

When we want to make comparison between groups of numbers it is good to ***have a single value*** that is considered to be a ***good representative*** of each group. This single value is called the **average** of the group.

An average which is representative is called typical average and an average which is not representative and has only a theoretical value is called a descriptive average.

A typical average should have the following characteristics:

- It should be rigidly defined.
- It should be based on all observation under investigation.
- It should be as little as affected by extreme observations.
- It should be capable of further algebraic treatment.
- It should be as little as affected by fluctuations of sampling.
- It should be ease to calculate and simple to understand.

2.2 Types of Measures of Central Tendency

Measures of Central Tendency:-give us information about the location *of the center of the distribution of data values*. A single value that approximately describes the characteristics of the entire mass of data is called *measures of central tendency*.

There are several different measures of central tendency; each has its own advantages and disadvantages. Among those:

- Mean (Arithmetic, Weighted, Geometric and Harmonic)
- Mode
- Median
- Quantiles (Quartiles, deciles and percentiles)

The choice of these averages depends up on which best fit the property under

discussion.

2.2.1 Mean

The Arithmetic Mean

- Is defined as the sum of the magnitude of the items divided by the number of items. The mean of $X_1, X_2, X_3 \ldots X_n$ is denoted by A.M ,m or X and is given by:

  $= =$, where n is sample size

  ☞ If we take an entire population the mean is denoted by μ and is given by:

  ☞ $= =$, where N is population size

  - ✓ If X_1 occurs f_1 times
  - ✓ If X_2 occurs f_2 times
  - .
  - .
  - ✓ If Xn occurs fn times

Then the mean will be =, where k is the number of classes and =n

Example: Calculate the arithmetic mean of the sample of numbers of students in 10 classes:

50 42 48 60 58 54 50 42 50 42

$= = = = 49.6$

In this case there are three 42's, one 48, three 50's, one 54, one 58 and one 60. The number of times each number occurs is called its frequency and the frequency is usually denoted by f. The information in the sentence above can be written in a table, as follows.

Value, x_i	42	48	50	54	58	60
Frequency, f_i	3	1	3	1	1	1
$x_i f_i$	126	48	150	54	58	60

The formula for the arithmetic mean for data of this type is

$= =$

In this case we have:

$= = = = 49.6$

The mean numbers of students in ten classes is 50.

Arithmetic Mean for Grouped Data

If data are given in the shape of a continuous frequency distribution, then the mean is obtained as follows:

=, where X_i =the class mark of the i^{th} class and f_i = the frequency of the i^{th} class

Example: The following frequency table gives the height (in inches) of 100 students in a college.

Class Interval (CI)	60-62	62-64	64-66	66-68	68-70	70-72	Total
Frequency (f)	5	18	42	20	8	7	100

Calculate the mean

Solution:

The formula to be used for the mean is as follows:

=

Let us calculate these values and make a table for these values for the sake of convenience.

Class Interval (CI)	60-62	62-64	64-66	66-68	68-70	70-72	**Total**
Frequency (f)	5	18	42	20	8	7	**100**
Mid-Point ()	61	63	65	67	69	71	
	305	1134	2730	1340	552	497	**6558**

Substituting these values with = 100, we get

= = = = 65.58

The mean height of students is 65.58

Exercises:

1.Marks of 75 students are summarized in the following frequency distribution:

Marks	No. of students
40-44	7
45-49	10
50-54	22
55-59	F_4
60-64	F_5
65-69	6
70-74	3

If 20% of the students have marks between 55 and 59

I. Find the missing frequencies f4 and f5.

II. Find the mean.

Special properties of Arithmetic mean

1. The sum of the deviations of a set of items from their mean is always zero. i.e. That is, =0
2. The sum of the squared deviations of a set of items from their mean is the minimum. i.e. ,
3. If the mean of , is , then
   a) The mean of ± k, ± k ,..., ± k will be ± k

b) The mean of will be k .

4. If $\overline{X}_1$ is the mean of n_1 observations, if $\overline{X}_2$ is the mean of n_2 observations, … ,if $\overline{X}_k$ is the mean of n_k observation, then the mean of all the observation in all groups often called the combined mean is given by:

Example: If the mean of one class of 50 students are 30 and the mean of marks of another class of 100 students are 40. What is the mean of all 150 students?

Solution: based on the above formula it is (50*30 + 100*40)/(50 + 100) =36.7.

5. If a wrong figure has been used when calculating the mean the correct mean can be obtained without repeating the whole process using:

   Correct Mean =, where n is total number of observations.

Example: An average weight of 10 students was calculated to be 65. Latter it was discovered that one weight was misread as 40 instead of 80 kg. Calculate the correct average weight.

Solutions:

Correct Mean =

Correct Mean =65 +**= 65+4=69k.g.**

6. The effect of transforming original series on the mean.

a) If a constant *k* is added/ subtracted to/from every observation then the new mean will be *the old mean*± *k* respectively.

b) If every observations are multiplied by a constant *k* then the new mean will be *k*old mean*

Example:

1. The mean of n Tetracycline Capsules X_1, X_2, …,X_n are known to be 12 gm. New set of capsules of another drug are obtained by the linear transformation $Y_i = 2X_i - 0.5$ (i = 1, 2, …, n) then what will be the mean of the new set of capsules

   Solutions:

   New Mean = 2*old mean - 0.5 = 2*12- 0.5=23.5

2. The mean of a set of numbers is 500. a) If 10 is added to each of the numbers in the set, then what will be the mean of the new set? b) If each of the numbers in the set are multiplied by -5, then what will be the mean of the new set?

   Solutions:

   a New Mean = Old Mean + 10 = 500+10= 510

   b. New Mean= -5*Old Mean= -5*500 = -2500

Merits and Demerits of Arithmetic Mean

Merits:

- It is rigidly defined.
- It is based on all observation.
- It is suitable for further mathematical treatment.
- It is stable average, i.e. it is not affected by fluctuations of sampling to some extent.
- It is easy to calculate and simple to understand.

Demerits:

- It is affected by extreme observations.
- It cannot be used in the case of open end classes.
- It cannot be determined by the method of inspection.
- It cannot be used when dealing with qualitative characteristics, such as intelligence, honesty, beauty.
- It can be a number which does not exist in a serious.
- Sometimes it leads to wrong conclusion if the details of the data from which it is obtained are not available.
- It gives high weight to high extreme values and less weight to low extreme values.

Weighted Mean

When a proper importance is desired to be given to different data a weighted mean is appropriate. Weights are assigned to each item in proportion to its relative importance.

Let X_1, X_2, …X_n be the value of items of a series and W_1,W_2, …W_n their corresponding weights, then the weighted mean denoted is defined as:

=

Example:

A student obtained the following percentage in an examination:

Statistics 60, Biology 75, Mathematics 63, Physics 59, and chemistry 55. Find the

students weighted arithmetic mean if weights 1, 2, 1, 3, 3 respectively are allotted to the subjects.

Solutions:

= = = =61.5

The Geometric Mean

If the observed values are measured as **ratios**, **proportions** or **percentages** and the series of observations contains one or more unusually large values geometric mean gives a better measure of central tendency than other means.

- ☞ The geometric mean of a set of n observation is the nth root of their product.
- ☞ The geometric mean of X_1, X_2 ,X_3 …X_n is denoted by G.M and given by:

$$G.M = \sqrt[n]{X_1 * X_2 * ... * X_n}$$

☞ Taking the logarithms of both sides

$$\log(G.M) = \log(\sqrt[n]{X_1 * X_2 * ... * X_n}) = \log(X_1 * X_2 * ... * X_n)^{\frac{1}{n}}$$

$$\Rightarrow \log(G.M) = \frac{1}{n}\log(X_1 * X_2 * .... * X_n) = \frac{1}{n}(\log X_1 + \log X_2 + ... + \log X_n)$$

$$\Rightarrow \log(G.M) = \frac{1}{n}\sum_{i=1}^{n}\log X_i$$

$\Rightarrow$ The logarithm of the G.M of a set of observation is the arithmetic mean of their logarithm.

Example: Find the G.M of the numbers 2, 4, 8.

Solutions:

$$G.M = \sqrt[n]{X_1 * X_2 * ... * X_n} = \sqrt[3]{2*4*8} = \sqrt[3]{64} = 4$$

The Harmonic Mean

The harmonic mean of X_1, X_2 , X_3 …X_n is denoted by H.M and given by:

H.M = , is called simple harmonic mean.

In a case of frequency distribution:

H.M = , where n=

If observations X_1, X_2 ,..., X_n have weights W_1,W_2 ,..., W_n respectively, then their harmonic mean is given by

H.M = , this is called Weighted Harmonic Mean.

Remark: *The Harmonic Mean is useful and appropriate in finding average speeds and average rates.*

2.2.2 Mode

- Mode is a value which occurs most frequently in a set of values.
- The mode may not exist and even if it does exist, it may not be unique.
- In case of discrete distribution the value having the maximum frequency is the model value.

Examples:

1. Find the mode of 5, 3, 5, 8, 9 Mode =5
2. Find the mode of 8, 9, 9, 7, 8, 2, and 5. It is a bimodal Data: 8 and 9
3. Find the mode of 4, 12, 3, 6, and 7. No mode for this data.

The mode of a set of numbers $X_1, X_2, ..., X_n$ is usually denoted by .

Mode for Grouped data

If data are given in the shape of continuous frequency distribution, the mode is defined as:

$$w = \textit{the size of the modal class}$$
$$\Delta_1 = f_{mo} - f_1$$
$$\Delta_2 = f_{mo} - f_2$$
$$f_{mo} = \textit{frequency of the modal class}$$
$$f_1 = \textit{frequency of the class preceeding the modal class}$$
$$f_2 = \textit{frequency of the class following the modal class}$$

Note: *The modal class is a class with the highest frequency.*

Example: Calculate the mode for the frequency distribution of data

C.I	1 – 5	6 - 10	11 – 15	16 – 20	21 - 25	26 - 30	31 - 35	**Total**
Freq.	4	8	12	6	3	4	3	**40**

Solution: By inspection, the mode lies in the third class, where L =10.5, f_{mod}= 12, f_1=8, f_2=6, w = 5

Using the formula, the mode is:

= 10.5 + (12-8)*5/(12-8)+(12-5) = 12.5

Merits and Demerits of Mode

Merits:

- It is not affected by extreme observations.
- Easy to calculate and simple to understand.
- It can be calculated for distribution with open end class

Demerits:

- Often its value is not unique.
- It is not based on all observations
- Mode may not exist in the series.
- It is not suitable for further mathematical treatment.

2.2.3 Median

In a distribution, median is the value of the variable which divides it in to two equal parts. In an **ordered** series of data median is an observation lying exactly in the middle of the series. It is the middle most value in the sense that the number of values less than the median is equal to the number of values greater than it and denoted by.

If $X_1, X_2, \ldots, X_n$ be the observations, then the numbers arranged in ascending order will be $X_{[1]}, X_{[2]}, \ldots X_{[n]}$, where $X_{[i]}$ is i^{th} smallest value. i.e. $X_{[1]} < X_{[2]} < \ldots < X_{[n]}$

Median for ungrouped data

Example: Find the median for the following data.

a) -5 15 10 5 0 2 1 4 6 and 8
b) 5 2 2 3 1 8 4

Solution;

a. The data in ascending order is given by:

-5 0 1 2 4 5 6 8 10 15

n=10 ➔n is even. The two middle values are 5^{th} and 6^{th} observations. So the median is,

= value =

b. The data in ascending order is given by:

1 2 2 3 4 5 8

The middle value is the 4^{th} observation. So the median is 3.

Median for grouped data

If data are given in the shape of continuous frequency distribution, the median is defined as:

where: the lower class boundary of the median class;
w = the class width of the median class;
the frequency of the median class; and
= the less than cum. freq. corresponding to the class preceding the median class.

Remark: *The median class is the class with the smallest cumulative frequency (less than type) greater than or equal to.*

Example: Calculate the median for the following frequency distribution.

C.I	1 – 5	6 - 10	11 – 15	16 – 20	21 - 25	26 - 30	31 - 35	**Total**
Freq.	4	8	12	6	3	4	3	**40**

Solution: Construct the less than cumulative frequency distribution, then:

C.I	1 – 5	6 - 10	11 – 15	16 – 20	21 - 25	26 - 30	31 - 35	**Total**
Freq.	4	8	12	6	3	4	3	**40**
Cuml. Freq.	4	12	24	30	33	37	40	

Since n = 40, 40/2 = 20, and the smallest CF greater than or equal to 20 is 24; thus, the median class is the third class. And for this class, L = 10.5, w = 5, f_{med} =12, CF = 12. Then applying the formula, we get: $\tilde{x}$ =10.5+ (20-12)*5/12=13.8

Merits and Demerits of Median

Merits:

- ☞ Median is a positional average and hence not influenced by extreme observations.
- ☞ Can be calculated in the case of open end intervals.
- ☞ Median can be located even if the data are incomplete.

Demerits:

- ☞ It is not a good representative of data if the number of items is small.
- ☞ It is not amenable to further algebraic treatment.
- ☞ It is susceptible to sampling fluctuations.

The Relationship of the Mean, Median and Mode

Comparing the Mean, Median, and the Mode

- ✓ If the data is skewed ***–avoid the mean.***
- ✓ If there is high gap around the middle- ***avoid the median***.
- ✓ The median is resistant to the influence of extreme data values or outliers.
- ✓ The mode has an advantage over both the mean and the median when the data is categorical since it is not possible to calculate the mean or median for this type of data.
- ✓ Mean=Median = Mode for symmetrical distribution; mean, median and mode coincide.

2.3 Measures of Location (Quantiles)

Quantiles are a measure which divides a given set of data in to approximately equal subdivision and are obtained by the same procedure to that of median. They are averages of position (non-central tendency). Some of these are quartiles, deciles and percentiles.

Quartiles: are values which divide the data set in to approximately four equal parts, denoted by. The first quartile () is also called the lower quartile and the third quartile () is the upper quartile. The second quartile () is the median.

- **Quartiles for Individual series**:

Let be n ordered observations. The i^{th} quartile () is the value of the item corresponding with the $[i(n+1)/4]^{th}$ position, i = 1, 2, 3.

That is, after arranging the data in ascending order, Q_1, Q_2, & Q_3 are, obtained by:

, and .

- **Quartiles for discrete data arranged in a frequency distribution**:-

Arranged in a frequency distribution this case also, we will follow the same procedure as the median. That is, we construct the less than cumulative frequency distribution and apply the formula of quartile for individual series.

- **Quartiles for grouped continuous data**:-

For continuous data, use the following formula:

Where i = 1, 2, 3, and L, w, f_{Qi} and CF are defined in the same way as the median.

i.e. $Q_1 = L +$, $Q_2 = L +$ $Q_3 = L +$

Deciles: are values dividing the data approximately in to ten equal parts, denoted by

.

- **Deciles for Individual Series**:

Let $x_1, x_2 \ldots x_n$ be n ordered observations. The i^{th} decile is the value of the item corresponding with the $[i(n+1)/10]^{th}$ position, i = 1, 2, . . . ,9.

That is, after arranging the data in ascending order, $D_1, D_2, \ldots$ & D_9 are, obtained by:

, . . . and.

- **Deciles for Discrete data arranged in a frequency distribution**:-

Arranged in a frequency distribution this case also, we will follow the same procedure as the median. That is, we construct the less than cumulative frequency distribution and apply the formula of deciles for individual series.

- **Deciles for continuous data:**

Apply the following formula and follow the procedures of quartile for continuous data.
, i = 1, 2... 9.

Then define the symbols in similar ways as we did in the case of quartiles for continuous data.

Percentiles: are values which divide the data approximately in to one hundred equal parts, and denoted by

- **Percentiles for individual Series**:

Let $x_1, x_2 \ldots x_n$ be n ordered observations. The i^{th} percentile is the value of the item corresponding with the $[i(n+1)/100]^{th}$ position, i = 1, 2, . . . ,99.

That is, after arranging the data in ascending order, $P_1, P_2, \ldots,$ & P_{99} are, obtained by:
, . . . and .

- **Percentiles for Discrete data arranged in a frequency distribution**:-

Arranged in a frequency distribution this case also, we will follow the same procedure as the median. That is, we construct the less than cumulative frequency distribution and apply the formula of percentile for individual series.

- **Percentiles for continuous data:**

Apply the following formula

, i = 1, 2... 99.

Then define the symbols similar ways as we did in the case of quartiles or deciles for continuous data.

Interpretations

1. is the value below which (i×25)% of the observations in the series are found (w'e i = 1, 2,3). For instance means the value below which 75% of observations in the given series are found.
2. is the value below which (i×10)% of the observations in the series are found (where i = 1, 2,...,9). For instance is the value below which 40% of the values are found in the series.
3. is the value below which i percent of the total observations are found (where i =1, 2,3,..,99). For example 60 percent of the observations in a given series are below.

Example: Marks of 50 students out of 85 is given below. Based on the data find ,

Marks	46-50	51-55	56-60	61-65	66-70	71-75	76-80
f_i	4	8	15	5	9	5	4

Solution: - first find the class boundaries and cumulative frequency distributions.

Marks	46-50	51-55	56-60	61-65	66-70	71-75	76-80
class boundary	45.5-50.5	50.5-55.5	55.5-60.5	60.5-65.5	65.5-70.5	70.5-75.5	75.5-80.5
f_i	4	8	15	5	9	5	4
Cum. Frequency	4	12	27	32	41	46	50

Q_1 Measure of $(n/4)^{th}$ value = 12.5^{th} value which lies in group 55.5 – 60.5

Q_1 = L + = 55.5 + = 55.7

D_4 Measure of $(4n/10)^{th}$ value = 20^{th} value which lies in group 55.5 – 60.5.

D_4 = L + = 55.5 + = 58.2

P_7 Measure of $(7n/100)^{th}$ value = 3.5^{th} value which lies in group 45.5 – 50.5

P_7 = L + = 45.5 + = 49.875.

2.4. Measures of Dispersion (Variation)

The scatter or spread of items of a distribution is known as dispersion or variation. In other words the degree to which numerical data tend to spread about an average value is called dispersion or variation of the data. -Measures of dispersions are statistical measures which provide ways of measuring the extent in which data are dispersed or spread out.

Objectives of measuring Variation:

- To judge the reliability of measures of central tendency
- To control variability itself.
- To compare two or more groups of numbers in terms of their variability.
- To make further statistical analysis.

Absolute and Relative Measures of Dispersion

The measures of dispersion which are expressed in terms of the original unit of a series are termed as absolute measures. Such measures are not suitable for comparing the variability of two distributions which are expressed in different units of measurement and different average size. appropriate measure of central tendency and are thus pure numbers independent of the units Relative measures of dispersions are a ratio or percentage of a measure of absolute dispersion to an of measurement.

For comparing the variability of two distributions (even if they are measured in the same unit), we compute the relative measure of dispersion instead of absolute measures of dispersion.

2.4.1 Types of Measures of Dispersion

Various measures of dispersions are in use. The most commonly used measures of dispersions are:

1) Range and relative range
2) Quartile deviation and coefficient of Quartile deviation
3) Mean deviation and coefficient of Mean deviation
4) Standard deviation and coefficient of variation.

The Range (R)

The range is the largest score minus the smallest score. It is a quick and dirty measure of variability, although when a test is given back to students they very often wish to know the range of scores. Because the range is greatly affected by extreme scores, it may give a distorted picture of the scores. The following two distributions have the same range, 13, yet appear to differ greatly in the amount of variability.

Distribution 1: 32 35 36 36 37 38 40 42 42 43 43 45

Distribution 2: 32 32 33 33 33 34 34 34 34 34 35 45

For this reason, among others, the range is not the most important measure of variability.

$R = L - S$, L = *largest observation*

S = *smallest observation*

Range for grouped data:

If data are given in the shape of continuous frequency distribution, the range is computed as:

$R = UCL_k - LCL_1$, UCL_k *is upper class limit of the last class*

UCL_1 *is lower class limit of the first class*

This is sometimes expressed as:

$R = X_k - X_1$, X_k *is class mark of the last class*

X_1 *is class mark of the last class.*

Merits and Demerits of range

Merits:

- It is rigidly defined.
- It is easy to calculate and simple to understand.

Demerits:

- It is not based on all observation.
- It is highly affected by extreme observations.
- It is affected by fluctuation in sampling.
- It is not liable to further algebraic treatment.
- It cannot be computed in the case of open end distribution.
- It is very sensitive to the size of the sample

Relative Range (RR)

It is also sometimes called coefficient of range and given by:

For a **continuous grouped distribution:**

Example:

1. Find the relative range of the above two distribution.(exercise!)
2. If the range and relative range of a series are 4 and 0.25 respectively. Then what is the

value of:

a) Smallest observation

b) Largest observation

Solutions :(2)

- ☞ R= L-S= 4 ------(1)
- ☞ RR=0.25, L+S=1------ (2), *then solving (1) and (2) we get* L =10 & S=6.

Variance and Standard Deviation

Variance

The variance is the average of squared deviations from the mean. Recall that the sum of squared deviations is minimum only when taken from the mean.

a) Population Variance ()

If we divide the variation by the number of values in the population, we get something called the population variance.

For ungrouped data (individual series) for population data

, where is the population arithmetic mean

- **For discrete data arranged in FD & for continuous grouped data**

, where the class mark of the i^{th} class is, is the frequency of the i^{th} class and N=

b) Sample Variance ()

One would expect the sample variance to simply be the population variance with the population mean replaced by the sample mean. However, one of the major uses of statistics is to estimate the corresponding parameter. This formula has the problem that the estimated value isn't the same as the parameter. To counteract this, the sum of the squares of the deviations is divided by one less than the sample size, this is because to get unbiased estimator.

For frequency distribution

If the values x_i have frequencies f_i (i=1,2,…,m), then the sample variance is given by:

.

The Standard Deviation

The square root value of variance is called standard deviation. The square root must be taken to get the units back the same as the original data values.

- ✓
- ✓

The following steps are used to calculate the sample variance:

1. Find the arithmetic mean.

2. Find the difference between each observation and the mean.

3. Square these differences.

4. Sum the squared differences.

5. Since the data is a sample, divide the number (from step 4 above) by the number of observations minus one, i.e., n-1 (where n is equal to the number of observations in the data set).

Example: Find the sample variance and standard deviation of:

x_i	2	4	5	6	8
f_i	2	2	3	1	2

Solution: Prepare the following table:

$\mathbf{x_i}$	$\mathbf{f_i}$	$\mathbf{f_i x_i}$	$\mathbf{x_i^2}$	$\mathbf{f_i x_i^2}$
2	2	4	4	8
4	2	8	16	32
5	3	15	25	75
6	1	6	36	36
8	2	16	64	128
Sum	10	49		279

Thus, n=.

= .

Properties of Variance & Standard Deviation

1. If a constant is added to (or subtracted from) all the values, the variance remains the same; i.e., for any constant k, $V(x_i \pm k) = V(x_i)$.
2. If each and every value is multiplied by a non-zero constant (k), the standard deviation is multiplied by and the variance is multiplied by k^2; i.e., $V(kx_i) = k^2 V(x_i)$.

3. Both the variance and the standard deviation give more weight to extreme values and less to those which are near to the mean.
4. If the standard deviation of X_1, X_2 …..X_n *is S* then the standard deviation of
   a) X_1+k, X_2+ k+,....+X_n+ k will also be S
   b) kX_1, kX_2+,.....+kX_n would be S
   c) a+kX_1, a+kX_2+… +a+kX_n would be S

Exercise: Verify each of the above relationship, considering *k* and *a* as constants.

Examples:

1. The mean and standard deviation of n Tetracycline Capsules X_1, X_2,..., X_n are known to be 12gm and 3gm respectively. New set of capsules of another drug are obtained by the linear transformation $Y_i = 2X_i - 0.5$ (i = 1, 2,…, n) then what will be the standard deviation of the new set of capsules

2. The mean and the standard deviation of a set of numbers are respectively 500 and 10.
   a. If 10 is added to each of the numbers in the set, then what will be the variance and standard deviation of the new set?
   b. If each of the numbers in the set are multiplied by -5, then what will be the variance and standard deviation of the new set?

Solutions:

1. Using above the new standard deviation =S =2*3 =6

2.
   a) They will remain the same.
   b) New standard deviation= S =5*10= 50

Coefficient of Variation (C.V)

Is defined as the ratio of standard deviation to the mean usually expressed as percents

The distribution having less C.V is said to be less variable or more consistent.

Examples:

1. An analysis of the monthly wages paid (in Birr) to workers in two firms A and B belonging to the same industry gives the following results

Value	Firm A	Firm B
Mean wage	52.5	47.5
Median wage	50.5	45.5

Variance 100 121

In which firm A or B is there greater variability in individual wages?

Solutions:

Calculate coefficient of variation for both firms.

Since C.VA < C.VB, in firm B there is greater variability in individual wages.

2. The students of Mechanical and Civil departments took *Introduction to Statistics and probability* course. At the end of the semester, the following information was recorded.

Department	Mechanical	Civil
Mean score	85	65
Standard deviation	25	12

Compare the relative dispersions of the two departments' scores using the appropriate way.

Solution:

Interpretation: Since the CV of Mechanical Department students is greater than that of Civil Department students, we can say that there is more dispersion relative to the mean in the distribution of Mechanical students' scores compared with that of Civil students.

3. A meteorologist interested in the consistency of temperatures in three cities during a given week collected the following data. The temperatures for the five days of the week in the three cities were

City1 25 24 23 26 17
City2 22 21 24 22 20
City3 32 27 35 24 28

Which city have the most consistent temperature, based on these data? (Exercise)

Standard Scores (Z-Scores)

The Z-score is the number of standard deviations that a given value X is below or above the mean and values above the mean have positive z-scores and values below the mean have

negative Z-scores. The numerical value of the Z-score reflects because of this Z-score is also referred to as relative measure of relative standing.

- Scores are generally meaningless by themselves unless they are compared to the distribution or scores from some reference group.
- In addition to comparison the data sets it is useful to transform a given data sets in to a standard normal distribution.

Properties of the Z-score

- ✓ The sum of Z-scores is always zero.
- ✓ The mean of Z-score is zero.
- ✓ The variance and standard deviation of z-score are equal to one.

Z-score computed from the population

Z-score computed from the sample

Example: What is the Z-score for the value of 14 in the following sample data set?

3 8 6 14 4 12 7 10

Solution:

= 8, SD = 3.8173 thus, Z =.

∴ The data value of 14 is located 1.57 standard deviations above the mean 8 because the z-score is positive.

Example: Suppose that a student scored 66 in Statistics and 80 in Mathematics. The score of the summary of the courses is given below.

Course	***Average score***	Standard deviation of the score
Statistics	51	12
Mathematics	72	16

In which course did the student scored better as compared to his classmates?

Solution:

Z-score of student in Statistics:

Z-score of student in Mathematics:

From these two standard scores, we can conclude that the student has scored better in Statistics course relative to his classmates than in Mathematics course.

Exercise

1. Two groups of people were trained 100km race and tested to find out which group is faster to complete the race. For the two groups the following information was given:

Value	Group one	Group two
Mean	10.4 min	11.9 min
Stan.dev.	1.2 min	1.3 min

Relatively speaking:

   a. Which group is more consistent in its performance?
   b. Suppose a person A from group one take 9.2 minutes while person B from Group two take 9.3 minutes, who was faster in completing the race? Why?