# **Chapter 3.**

**3-2 Measures of center**

**Statistic:** is a characteristic or measure obtained by using the data values from a sample.

**Parameter:** is a characteristic or measure obtained by using all the data values for a specific population.

***Average*** means the center of the distribution or the most typical case. Several measures can be used for an average. Measures of average are also called *measures of central tendency* and include the *mean, median, mode, and, midrange.*

**Mean:** The mean is the sum of the values, divided by the total number of values. The symbol represents the ***sample mean.***



where n represents the total number of values in the sample.

For a ***population mean***, the Greek letter **μ** (mu) is used.



where N represents the total number of values in the population.

**Rounding Rule for the Mean:** The mean should be rounded to one more decimal place than occurs in the raw data. For example if the raw data are given in whole numbers, the mean should be rounded to the nearest tenth. If the data are given in tenths, the mean should be rounded to the nearest hundredth, and so on.

**Finding the mean for group data:**

Step1. Make a table as shown.

A B C D

Class Frequency (f) Midpoint (Xm) f • Xm

Step2. Find the midpoints of each class and place them in

column C.

Step3. Multiply the frequency by the midpoint for each

class and place the product in column D.

Step4. Find the sum of column D.

Step5. Divide the sum obtained in column D by the sum of

the frequencies obtained in column B.

The formula for mean is



**The Median:**

When the data set is ordered, it is called a ***data array.***

The **median** is the midpoint of the data array. The symbol for the median is MD.

\*When there is an even number of values in the data set, the median will fall between two given values.

**Computing the median of a data array:**

Step1. Arrange the data in order.

Step2. Select the middle point.

**The Mode:** The value that occurs most often in a data set.

\*A data set can have more than one mode or no mode at all.

\*If each value occurs only once, there is no mode.

\*The mode for grouped data is the modal class. The ***modal class*** is the class with the largest frequency.

**The midrange:** is the sum of the lowest and highest values in the data set, divided by 2. The midrange is denoted by MR.



**Distribution Shapes:** The three most important shapes are positively skewed, symmetric, and negatively skewed.

***1). Positively skewed or right skewed distribution*:** the majority of the data values fall to the left of the mean and cluster at the lower end of the distribution; the “tail” is to the right.

The mean is to the right of the median, and the mode is to the left of the median.

***2). Symmetric distribution*:** The data values are evenly distributed on both sides of the mean.

When the distribution is unimodal, the mean, median, and mode are the same and are at the center of the distribution.

e.g. Heights of adult males.

***3). Negatively skewed or left-skewed*:** When the majority of the data values fall to the right of the mean and cluster at the upper end of the distribution, with the tail to the left.

Mean is to the left of the median, and the mode is to the right of the median.

### 3-3 Measures of Variation

For the spread or variability of a data set, three measures are commonly used: *range, variance, and standard deviation.*

**Range:** R = highest value – lowest value

\*If the range of data is a large number, to have a more meaningful statistic to measure the variability, statisticians use measures called the variance and standard deviation.

***Population* Variance and Standard Deviation:**

The **variance** is the average of the squares of the distance each value is from the mean. Variance is denoted by σ2 (σ is Greek lowercase letter sigma).



Where X = individual value

μ = Population mean

N = population size

The **Standard deviation** is the square root of the variance. The symbol for population standard deviation is σ.



***Sample* Variance and Standard Deviation:**

Variance = 

Where  = sample mean

n = sample size

Standard deviation = s =

**Shortcut or computational formulas for s2 and s:**

**Variance: **

**Standard Deviation: **

**Procedure for finding the sample variance and standard deviation for *Grouped data*:**

# Step1. Make a table as shown and find the midpoint of each

# class.

**A B C D E**

**Class Frequency Midpoint f•Xm f•Xm2**

fXm

Step2. Multiply the frequency by the midpoint for each class and place the products in column D.

Step3. Multiply the frequency by the square of the midpoint and place the products in column E.

Step4. Find the sums of column B, D, and E. The sum of column B is n. The sum of column D is ∑f • Xm. The sum of column E is ∑f • Xm2.

Step5. Substitute in the formula and solve to get the standard deviation



**Uses of variance and standard deviation:**

# Variance and standard deviation can be used

1. to determine the spread of the data.
2. to determine the consistency of a variable. e.g. fitting of nuts and bolt.

# 3. to determine the number of data values that fall within a specified interval in a distribution.

4. in inferential statistics.

**Chebyshev’s Theorem:** specifies the proportions of the spread in terms of standard deviation.

Chebyshev’s Theorem applies to any distribution regardless of its shape. When the distribution is bell shaped, it constitutes The Empirical Rule.

**The Empirical (Normal) Rule:**

When the distribution is *bell-shaped (or normal*),

1) Approximately 68% of the data values will fall within 1 standard deviation of the mean.

2) Approximately 95% of the data values will fall within 2 standard deviation of the mean.

3) Approximately 99.7% of the data values will fall within 3 standard deviation of the mean.

**3-4 Measures of Position:** are standard score or z score, percentiles, deciles and quartiles. These are used to find the relative position of a data value in the data set.

**Z score:** It tells the number of standard deviations that a data value falls above or below the mean.



*\*If the z score is positive, the score is above the mean.*

*If the z score is negative, the score is below the mean.*

*If the z score is 0, the score is same as the mean.*

*\* Higher the z score, better the relative position of the data.*

**Percentiles:** Percentiles divide the data set into 100 equal groups.