2.1 The Nature of Statistics

Statistics refer to the methods used to collect and analyze data. Statistical methods include descriptive statistics and statistical inference (inferential statistics).

- **Descriptive statistics:** It describes the properties of a large data set by summarizing it in an effective manner.
- **Statistical inference:** It involves use of a sample to make forecasts, estimates, or judgments about the characteristics of a population

2.2 Populations and Samples

- A **population** is a complete set of outcomes or all members of a specified group.
- A **parameter** describes a characteristic of a population e.g. mean value, the range of investment returns, and the variance.

Since analyzing the entire population involves high costs, it is preferred to use a sample.

- A **sample** is a subset of a population.
- A **sample statistic** or **statistic** describes a characteristic of a sample.

2.3 Measurement Scales

Measurement scales are the specific set of rules used to assign a symbol to the event in question. There are four types of measurement scales.

a) Nominal Scale: It is a simple classification system under which the data is categorized into various types.

- It does not rank the data.
- It is the weakest level of measurement.

Example:

Mutual funds can be categorized according to their investment strategies i.e.

- Mutual Fund 1 refers to a small-cap value fund.
- Mutual Fund 2 refers to a large-cap value fund.

b) Ordinal Scale: This scale categorizes data into various categories and also rank them into an order based on some characteristics.

- It is a stronger level of measurement relative to nominal scale.
- However, the intervals separating the ranks in ordinal scale cannot be compared with each other.

Example:

Under Morningstar and Standard & Poor's star ratings for mutual funds,

- A fund that is assigned 1 star represents a fund with relatively poor performance.
- A fund that is assigned 5 stars represents a fund with relatively superior performance.

c) Interval Scale: This scale rank the data into an order based on some characteristics and the differences between scale values are ***equal*** e.g. Celsius and Fahrenheit scales.

- The zero point of an interval scale does not reflect a true zero point or natural zero e.g. 0°C does not represent absence of temperature; rather, it reflects a freezing point of water.
- As a result, it cannot be used to compute ratios e.g. 40°C is two times larger than 20°C; however, it does not represent two times as much temperature.
- Since difference between scale values are equal, scale values can be added and subtracted meaningfully.

Example:

The difference in temperature between 15°C and 20°C is the same amount as the difference between 40°C and 45°C. Also, 10°C + 5°C = 15°C

d) Ratio Scale: It is the strongest level of measurement. Under this scale,

- The data is ranked based on some characteristics.
- The differences between scale values are equal; therefore, scale values can be added and subtracted meaningfully.
- A true zero point as the origin exists. E.g. zero money means no money.
  - o Thus, it can be used to compute ratios and to add and subtract amounts within the scale.

Example:

Money is measured on a ratio scale i.e. the purchasing power of $100 is twice as much as that of $50.

Practice: Example 1, Volume 1, Reading 8.

3. SUMMARIZING DATA USING FREQUENCY DISTRIBUTIONS

Data can be summarized using a frequency distribution. In a **Frequency distribution,** data is grouped into ***mutually exclusive*** categories and shows the number of observations in each class.

- It is also useful to identify the shape of the distribution.

Construction of a Frequency Distribution table:

Step 1: *Arrange the data in ascending order.*
Step 2: *Calculate the* ***range*** *of the data.*
Range = Maximum Value - Minimum value
Step 3: *Choose the appropriate number of classes (k):* Determining the number of classes involves judgment.

NOTE:

A large value of k is useful to obtain detailed information regarding the extreme values of a distribution.

Step 4: *Determine the class interval or width using the following formula i.e.*

$$i \geq (H-L)/k$$

where,

i= Class interval
H = Highest observed value
L = Lowest observed value
k= Number of classes

Interval: An interval represents a set of values within which an observation lies.

- If too few intervals are used, then the data is over-summarized and may ignore important characteristics.
- If too many intervals are used, then the data is under-summarized.
- The smaller (greater) the value of k, the larger (smaller) the interval.

Example:

Suppose,

H = \$35,925
L = \$15,546
k= 7

Class interval = (\$35,925 - \$15,546)/7 = \$2,911≈ \$3,000.

It is important to note that:

- We will always round up (not down), to ensure that the final class interval includes the maximum value of the data.
- The class intervals (also known as ranges or bins) do not overlap.

Step 5: *Set the individual class limits* i.e.

- Ending points of intervals are determined by successively adding the interval width to the *minimum* value.
- The last interval would be the one, which includes the *maximum* value.

NOTE:

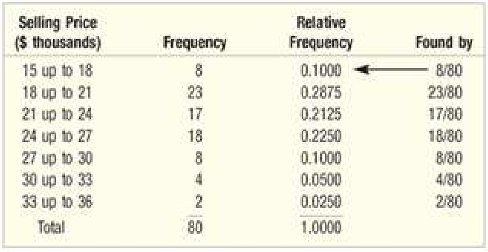
The notation (20,000 to 25,000) means 20,000 ≤ observation < 25,000 ➔A square bracket shows that the endpoint is included in the interval.

Step 6: *Count the number of observations in each class interval.*

Absolute Frequency: The actual number of observations in a given class interval is called the absolute frequency or simply frequency; as shown in the table below i.e. there are 8 observations that fall under the price interval 15 up to 18.

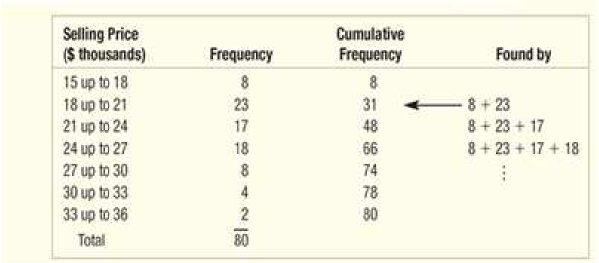
Relative frequency:

Relative frequency = Absolute frequency / Total number of observations



Selling Price ($ thousands)	Frequency	Relative Frequency	Found by
15 up to 18	8	0.1000	8/80
18 up to 21	23	0.2875	23/80
21 up to 24	17	0.2125	17/80
24 up to 27	18	0.2250	18/80
27 up to 30	8	0.1000	8/80
30 up to 33	4	0.0500	4/80
33 up to 36	2	0.0250	2/80
Total	80	1.0000	

Cumulative Absolute Frequency: The cumulative *absolute* frequency is found by adding up the absolute frequencies. It reflects the ***number*** of observations that are less than the upper limit of each interval.



Selling Price ($ thousands)	Frequency	Cumulative Frequency	Found by
15 up to 18	8	8	
18 up to 21	23	31	8 + 23
21 up to 24	17	48	8 + 23 + 17
24 up to 27	18	66	8 + 23 + 17 + 18
27 up to 30	8	74	⋮
30 up to 33	4	78	
33 up to 36	2	80	
Total	80		

Cumulative Relative Frequency: The cumulative *relative* frequency is found by adding up the relative frequencies. It reflects the ***percentage*** of observations that are less than the upper limit of each interval.

E.g. in the table above after the "***relative frequency***", the cumulative relative frequency for the

- 2nd class interval would be 0.10 + 0.2875 = 0.3875 ➔ it indicates that 38.75% of the observations lie below the selling price of 21.
- 3rd class interval would be 0.3875 + 0.2125 = 0.60 ➔ it indicates that 60% of the observations lie below the selling price of 24.

E.g. in the table below cumulative relative frequency for the 2nd class interval would be 0.10 + 0.2875 = 0.3875 and for the 3rd class interval would be 0.3875 + 0.2125 = 0.60

NOTE:

The frequency distributions of annual returns cannot be compared directly with the frequency distributions of monthly returns.

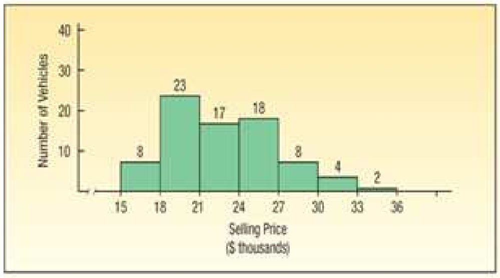
For details, refer to discussion before table 4, Volume 1, Reading 8.

Practice: Example 2, Volume 1, Reading 8.

4.1 The Histogram

A histogram is the graphical representation of a frequency distribution.
s

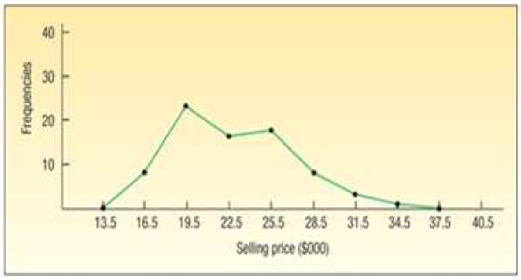
- The classes are plotted on the *horizontal* axis.
- The class frequencies are plotted on the *vertical* axis.
- The heights of the bars of histogram represent the absolute class frequencies.
- Since the classes have no gaps between them, there would be no gaps between the bars of the histogram as well.



4.2 The Frequency Polygon and the Cumulative Frequency Distribution

Frequency polygon: It also graphically represents the frequency distribution.

- The mid-point of each class interval is plotted on the *horizontal* axis.
- The corresponding absolute frequency of the class interval is plotted on the *vertical* axis.
- The points representing the intersections of the class midpoints and class frequencies, are connected by a line.

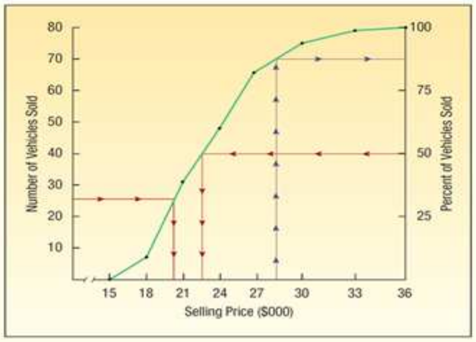


Cumulative frequency distribution: This graph can be used to determine the number or the percentage of the observations lying between a certain values. In this graph,

- Cumulative absolute or cumulative relative frequency is plotted on the *vertical* axis.
- The upper interval limit of the corresponding class interval is plotted on the *horizontal* axis.
  - o For extreme values (both negative and positive), the cumulative distribution tends to flatten out.
  - o Steeper (flatter) slope of the curve indicates large (small) frequencies (# of observations).

NOTE:

Change in the cumulative relative frequency = Relative frequency of the next interval.



5. MEASURES OF CENTRAL TENDENCY

A measure of central tendency indicates the center of the data. The most commonly used measures of central tendency are:

1. **Arithmetic mean or Mean:** It is the sum of the observations in the dataset divided by the number of observations in the dataset.

2. **Median:** It is the middle number when the observations are arranged in ascending or descending order. A given frequency distribution has only one median.

3. **Mode:** It is the observation that occurs most frequently in the distribution. Unlike median, a mode is not unique which implies that a distribution may have more than one mode or even no mode at all.

4. **Weighted mean:** It is the arithmetic mean in which observations are assigned different weights. It is computed as:

$$\bar{X}_w = \sum_{i=1}^{n} w_i X_i = (w_1X_1 + w_2X_2 + \cdots + w_nX_n)$$

where,

$X_1, X_2, \ldots, X_n$ = observed values
$w_1, w_2, \ldots, w_3$ = Corresponding weights, sum to 1.

- An arithmetic mean is a special case of weighted mean where all observations are equally weighted by the factor 1/ n (or I/N).
- A positive weight represents a long position and a negative weight represents a short position.
- **Expected value:** When a weighted mean is computed for a forward-looking data, it is referred to as the expected value.

Example:

Weight of stocks in a portfolio = 0.60
Weight of bonds in a portfolio = 0.40
Return on stocks = –1.6%
Return on bonds = 9.1%

A portfolio's return is the weighted average of the returns on the assets in the portfolio i.e.

Portfolio return = (w_{stock} × R_{stock}) + (w_{bonds} × R_{bonds})
= 0.60(-1.6%) + 0.40 (9.1%) = 2.7%.

Practice: Example 6, Volume 1, Reading 8.

5. **Geometric mean (GM):** The geometric mean can be used to compute the mean value over time to compute the growth rate of a variable.

$$G = \sqrt[n]{X_1X_2X_3 \ldots X_n}$$

with Xi ≥ 0 for i = 1, 2, …, n.

Or

$$In\ G = \frac{1}{n}\, In(X_1X_2X_3 \ldots X_n)$$

or as

$$In\ G = \frac{\sum_{i=1}^{n} InX_n}{n}$$

$$G = e^{lnG}$$

- It should be noted that the geometric mean can be computed only when the product under the radical sign is non-negative.

The geometric mean return over the time period can be computed as:

$$R_{Geom} = [(1+R_1)(1+R_2)\ldots(1+R_n)]^{1/n} - 1$$

- Geometric mean returns are also known as compound returns.

Advantages of Measures of central tendency:

- Widely recognized.
- Easy to compute.
- Easy to apply.

5.1.1) The Population Mean

It is the arithmetic mean of the total population and is computed as follows:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

where,

μ = Population mean
N = Number of observations in the entire population
X_i = i^{th} observation.

- The population mean is a population parameter.
- A given population has only one mean.

5.1.2) The Sample Mean

The sample mean is the arithmetic mean value of a sample; it is computed as:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

where,

$\bar{X}$ = sample mean
X_i = i^{th} observation
n = number of observations in the sample

- The sample mean is a statistic.
- It is not unique i.e. for a given population; different samples may have different means.

Cross-sectional mean: The mean of the cross-sectional data i.e. observations at a specific point in time is called cross-sectional mean.

Time-series mean: The mean of the time-series data e.g. monthly returns for the past 10 years is called time-series mean.

Practice: Example 3, Volume 1, Reading 8.

5.1.3) Properties of the Arithmetic Mean

Property 1: The sum of the deviations* around the mean is always equal to 0.

*The difference between each outcome and the mean is called a deviation.

Property 2: The arithmetic mean is sensitive to extreme values i.e. it can be biased upward or downward by extremely large or small observations, respectively.

Advantages of Arithmetic Mean:

- The mean uses all the information regarding the size and magnitude of the observations.
- The mean is also easy to calculate.
- Easy to work with algebraically

Limitation: The arithmetic mean is highly affected by outliers (extreme values).

- **Trimmed Mean:** It is the arithmetic mean of the distribution computed after excluding a stated small % of the lowest and highest values.
- **Winsorized mean:** In a winsorized mean, a stated % of the lowest values is assigned a specified low value and a stated % of the highest values is assigned a specified high value and then a mean is computed from the restated data. E.g. in a 95% winsorized mean,
  - o The bottom 2.5 % of values are set = 2.5th percentile value.
  - o The upper 2.5% of values are set = 97.5th percentile value.

5.2 The Median

Population median: A population median divides a population in half.

Sample median: A sample median divides a sample in half.

Steps to compute the Median:

1. Arrange all observations in ascending order i.e. from the smallest to the largest.
2. When the number of observations (n) is odd, the median is the center observation in the ordered list i.e.

   Median will be located at = $\frac{(n+1)}{2}$ position

- (n+1)/2 only identifies the location of the median, not the median itself.

3. When the number of observations (n) is even, then median is the mean of the two center observations in the ordered list i.e.

   Median will be located at mean of $\frac{n}{2}$ and $\frac{(n+1)}{2}$.

Advantage: Median is not affected by extreme observations (outliers).

Limitations:

- It is time consuming to calculate median.
- The median is difficult to compute.
- It does not use all the information about the size and magnitude of the observations.
- It only focuses on the relative position of the ranked observations.

Example:

Suppose, current P/Es of three firms are 16.73, 22.02, and 29.30.

n = 3 → (n + 1) / 2 = 4/ 2 = 2nd position.

Thus, the median P/E is 22.02.

Practice: Example 4, Volume 1, Reading 8.

5.3 The Mode

Population mode: A population mode is the most frequently occurring value in the population.

Sample mode: A sample mode is the most frequently occurring value in the sample.

Unimodal Distribution: A distribution that has only one mode is called a unimodal distribution.

Bimodal Distribution: A distribution that has two modes is called a bimodal distribution.

Trimodal Distribution: A distribution that has three modes is called a Trimodal distribution.

A distribution would have no mode when all the values in a data set are different.

Modal Interval: Data with continuous distribution (e.g. stock returns) may not have a modal outcome. In such cases, a modal interval is found i.e. an interval with the largest number of observations (highest frequency). The modal interval always has the highest bar in the histogram.

Important to note: The mode is the only measure of central tendency that can be used with nominal data.

Practice: Example 5, Volume 1, Reading 8.

5.4.2) The Geometric Mean

Geometric mean v/s Arithmetic mean:

- The geometric mean return represents the growth rate or compound rate of return on an investment.
- The arithmetic mean return represents an average single-period return on an investment.
- The geometric mean is *always* ≤ arithmetic mean.
- When there is no variability in the observations (i.e. when all the observations in the series are the same), geometric mean = arithmetic mean
- The greater the variability of returns over time, the more the geometric mean will be lower than the arithmetic mean.
- The geometric mean return decreases with an increase in standard deviation (holding the arithmetic mean return constant).
- In addition, the geometric mean ranks the two funds differently from that of an arithmetic mean.

Practice: Example 7 & 8, Volume 1, Reading 8.

5.4.3) The Harmonic Mean

$$Harmonic\ Mean(H.M)\bar{X}_H = n/\sum_{i=1}^{n}(\frac{1}{X_i})$$

with $X_i > 0$ for i = 1,2, …, n.

- It is a special case of the weighted mean in which each observation's weight is inversely proportional to its magnitude.

Practice: Example on 5.4.3, Volume 1, Reading 8.

Important to note:

- Harmonic mean formula cannot be used to compute average price paid when different amounts of money are invested at each date.
- When all the observations in the data set are the same, geometric mean = arithmetic mean = harmonic mean.
- When there is variability in the observations, harmonic mean < geometric mean < arithmetic mean.

6. OTHER MEASURES OF LOCATION: QUANTILE

Measures of location: Measures of location indicate both the center of the data and location or distribution of the data. Measures of location include measures of central tendency and the following four measures of location:

- Quartiles
- Quintiles
- Deciles
- Percentiles

Collectively these are called "***Quantiles"***.

6.1 Quartiles, Quintiles, Deciles, and Percentiles

1) **Quartiles** divide the distribution into four different parts.

- First Quartile = Q1 = 25th percentile i.e. 25% of the observations lie at or below it.
- Second Quartile = Q2 = 50th percentile i.e. 50% of the

observations lie at or below it.

- Third Quartile = Q3 = 75th percentile i.e. 75% of the observations lie at or below it.

2) Quintiles divide the distribution into five different parts. In terms of percentiles, they can be specified as P_{20}, P_{40}, P_{60}, & P_{80}.

3) Deciles divide the distribution into ten different parts.

4) Percentiles divide the distribution into hundred different parts. The position of a percentile in an array with n entries arranged in ascending order is determined as follows:

$$L_y = (n+1)\frac{y}{100}$$

where,

y = % point at which the distribution is being divided.
Ly = location (L) of the percentile (Py).
n = number of observations.

- The larger the sample size, the more accurate the calculation of percentile location.

Example:

Dividend Yields on the components of the DJ Euros STOXX 50

No.	Company	Dividend Yield(%)
1	AstraZeneca	0.00
2	BP	0.00
3	Deutsche Telekom	0.00
4	HSBC Holdings	0.00
5	Credit Suisse Group	0.26
6	L'Oreal	1.09
7	SwissRe	1.27
8	Roche Holding	1.33
9	Munich Re Group	1.36
10	General Assicurazioni	1.39
11	Vodafone Group	1.41
12	Carrefour	1.51
13	Nokia	1.75
14	Novartis	1.81
15	Allianz	1.92
16	Koninklije Philips Electronics	2.01
17	Siemens	2.16
18	Deutsche Bank	2.27
19	Telecom Italia	2.27
20	AXA	2.39
21	Telefonica	2.49
22	Nestle	2.55
23	Royal Bank of Scotland Group	2.60
24	ABN-AMRO Holding	2.65
25	BNP Paribas	2.65
26	UBS	2.65
27	Tesco	2.95
28	Total	3.11
29	GlaxoSmithKline	3.31
30	BT Group	3.34
31	Unilever	3.53
32	BASF	3.59
33	Santander Central Hispano	3.66
34	Banco Bilbao VizcayaArgentaria	3.67
35	Diageo	3.68
36	HBOS	3.78
37	E.ON	3.87
38	Shell Transport and Co.	3.88
39	Barclays	4.06
40	Royal Dutch Petroleum Co.	4.27
41	Fortus	4.28
42	Bayer	4.45
43	DiamlerChrysler	4.68
44	Suez	5.13
45	Aviva	5.15
46	Eni	5.66
47	ING Group	6.16
48	Prudential	6.43
49	Lloyds TSB	7.68
50	AEGON	8.14

Source: Example 9, Table 17, Volume 1, Reading 8.s

Calculating 10th percentile (P_{10}): Total number of observations in the table above = n = 50

$$L_{10} = (50 + 1) \times (10 / 100) = 5.1$$

- It implies that 10th percentile lies between 5th observation (X_5 = 0.26) and 6th observation (X_6 = 1.09).

Thus,
$P_{10} = X_5 + (5.1 – 5) (X_6 – X_5) = 0.26 + 0.1 (1.09 – 0.26)$
$= 0.34\%$

Calculating 90th percentile (P_{90}):

$$L_{90} = (50 + 1) \times (90 / 100) = 45.9$$

- It implies that 90th percentile lies between the 45th observation ($X_{45} = 5.15$) and 46th observation ($X_{46} = 5.66$).

Thus,
$P_{90} = X_{45} + (45.9 - 45) (X_{46} - X_{45}) = 5.15 + 0.90 (5.66 - 5.15) = 5.61\%$

Calculating 1stQuartile (i.e.P_{25}):

$$L_{25} = (50 + 1) \times (25 / 100) = 12.75$$

- It implies that 25th percentile lies between the 12th observation ($X_{12} = 1.51$) and 13th observation ($X_{13} = 1.75$).

Thus,

$$P_{25} = Q_1 = X_{12} + (12.75 - 12) (X_{13} - X_{12}) = 1.51 + 0.75 (1.75 - 1.51) = 1.69\%$$

Calculating 2nd Quartile (i.e.P_{50}):

$$L_{50} = (50 + 1) \times (50 / 100) = 25.5$$

- It implies that P_{50} lies between the 25th observation ($X_{25} = 2.65$) and 26th observation ($X_{26} = 2.65$).
- Since, $X_{25} = X_{26} = 2.65$, no interpolation is needed.

Thus,

$P_{50} = Q2 = 2.65\% =$ Median

Calculating 3rd Quartile (i.e.P_{75}):

$$L_{75} = (50 + 1) \times (75 / 100) = 38.25$$

- It implies that P_{75} lies between the 38th observation ($X_{38} = 3.88$) and 39th observation ($X_{39} = 4.06$).

Thus,

$$\begin{aligned} P_{75} = Q3 &= X_{38} + (38.25 - 38) (X_{39} - X_{38}) \\ &= 3.88 + 0.25 (4.06 - 3.88) \\ &= 3.93\% \end{aligned}$$

Calculating 20th percentile (P_{20}) = 1st Quintile:

$$L_{20} = (50 +1) \times (20 /100) = 10.2$$

- It implies that P_{20} lies between the 10th observation ($X_{10} = 1.39$) and 11th observation ($X_{11} = 1.41$).

Thus,

1st quintile = $P_{20} = X_{10} + (10.2 - 10) (X_{11} - X_{10}) = 1.39 + 0.20 (1.41 - 1.39) = 1.394\%$ or 1.39%

Source: Example 9, Volume 1, Reading 8.

6.2 Quantiles in Investment Practice

Quantiles are frequently used by investment analysts to rank performance i.e. portfolio performance. For example, an analyst may rank the portfolio of companies based on their market values to compare performance of small companies with large ones i.e.

- 1st decile contains the portfolio of companies with the smallest market values.
- 10th decile contains the portfolio of companies with the largest market values.

Quantiles are also used for investment research purposes.

7. MEASURES OF DISPERSION

The variability around the central mean is called Dispersion. The measures of dispersion provide information regarding the **spread** or **variability** of the data values.

Relative dispersion: It refers to the amount of dispersion/variation relative to a reference value or benchmark e.g. coefficient of variation. (It is discussed below).

Absolute Dispersion: It refers to the variation around the mean value without comparison to any reference point or benchmark. Measures of absolute dispersion include:

1) Range:

Range = Maximum value - Minimum value

Advantage: It is easy to compute.

Disadvantages:

- It does not provide information regarding the shape of the distribution of data.
- It only reflects extremely large or small outcomes that may not be representative of the distribution.

NOTE:

Interquartile range (IQR) = Third quartile - First quartile
= Q3 – Q1

- It reflects the length of the interval that contains the middle 50% of the data.
- The larger the interquartile range, the greater the dispersion, all else constant.

2) Mean absolute deviation (MAD): It is the average of the ***absolute*** values of deviations from the mean.

$$MAD = \frac{\sum_{i=1}^{n}|X_t - \bar{X}|}{n}$$

where,

$\bar{X}$ = Sample mean
n = Number of observations in the sample

- The greater the MAD, the riskier the asset.

Example:

Suppose, there are 4 observations i.e. 15, -5, 12, 22.

Mean = (15 – 5 + 12 + 22)/4 = 11%
MAD = (| 15 – 11| + |–5 – 11| + |12 – 11| + |22 – 11|)/4
= 32/4 = 8%

Advantage:

MAD is superior relative to range because it is based on all the observations in the sample.

Drawback:

MAD is difficult to compute relative to range.

3) Variance: Variance is the average of the squared deviations around the mean.

4) Standard deviation (S.D.): Standard deviation is the positive square root of the variance. It is easy to interpret relative to variance because standard deviation is expressed in the same unit of measurement as the observations.

7.3.1) Population Variance

The population variance is computed as:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

where,

μ= Population mean
N = Size of the population

Example:

Returns on 4 stocks: 15%, –5%, 12%, 22%
Population Mean (μ) = 11%

$$\sigma^2 = \frac{(15-11)^2 + (-5-11)^2 + (12-11)^2 + (22-11)^2}{4}$$
$$= 98.5$$

7.3.2) Population Standard Deviation

It is computed as:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$$

$$Standard\ deviation(\sigma) = \sqrt{98.5} = 9.9\%$$

7.4.1) Sample Variance

It is computed as:

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

where,

$\bar{X}$=Sample mean
n = Number of observations in the sample

- The sample mean is defined as an ***unbiased estimator*** of the population mean.
- **(n – 1)** is known as the number of degrees of freedom in estimating the population variance.

7.4.2) Sample Standard Deviation

It is computed as:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

Important to note:

- The MAD will always be ≤ S.D. because the S.D. gives more weight to large deviations than to small ones.
- When a constant amount is added to each observation, S.D. and variance remain unchanged.

Practice: Example 10, 11 & 12, Volume 1, Reading 8.

7.5 Semivariance, Semideviation, and Related Concepts

Semivariance is the average squared deviation ***below*** the mean.

$$\sum_{For\ all\ X_i \le \bar{X}} (X_i - \bar{X})^2/(n-1)$$

Semi-deviation (or semi-standard deviation) is the positive square root of semivariance.

- Semi-deviation will be < Standard deviation because standard deviation overstates risk.

Example:

Returns (in %): 16.2, 20.3,9.3, -11.1, and -17.0.

Thus, n = 5

Mean return = 3.54%

Two returns, -11.1 and -17.0, are < 3.54%.

Semi-variance =((-11.1 - 3.54)^2 + (-17.0- 3.54)^2) / 5 – 1
=636.2212/4 = 159.0553

Semi-deviation= $\sqrt{159.0553}$ = 12.6%.

Target semi-variance is the average squared deviation below a ***stated target***.

$$\sum_{For\ all\ X_i \leq B} (X_i - B)^2/(n-1)$$

where,

B = target value,
n = number of observations.

Target semi-deviation is the positive square root of the target semi-variance.

NOTE:

- Semivariance (or Semideviation) and target Semivariance (or target Semideviation) are difficult to compute compared to variance.
- For symmetric distributions, semi-variance = variance.

Example:

Stock returns = 16.2, 20.3, 9.3%, –11.1% and –17.0%.
Target return = B = 10%

Target semi-variance = ((9.3 –10.0)^2 + (–11.1 – 10.0)^2 + (–17.0 – 10.0)^2)/(5 – 1)
= 293.675

Target semi-deviation = $\sqrt{293.675}$ = 17.14%

7.6 Chebyshev's Inequality

Chebyshev's inequality can be used to determine the ***minimum*** % of observations that must fall within a given interval around the mean; however, it does not give any information regarding the maximum % of observations.

According to Chebyshev's inequality:

*The proportion of any set of data lying within **k** standard deviations of the mean is always at least (1 – 1/ (K^2))➔ for all k >1.*

Regardless of the shape of the distribution and for samples and populations and for discrete and continuous data:

- Two S.D. interval around the mean ***must*** contain at least 75% of the observations.
- Three S.D. interval around the mean ***must*** contain at least 89% of the observations.

Example:

When k = 1.25, then according to Chebyshev's inequality,

- The minimum proportion of the observations that lie within + 1.25s is (1 - 1/ (1.25)^2) = 1 - 0.64 = 0.36 or 36%.

Practice: Example 13, Volume 1, Reading 8.

7.7 Coefficient of Variation

Coefficient of Variation (CV) measures the amount of risk (S.D.) per unit of mean value.

$$CV = \left(\frac{S}{\bar{X}}\right)$$

When stated in %, CV is:

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

where,

s = sample S.D.
$\bar{X}$ *= sample mean.*

- CV is a scale-free measure (i.e. has no units of measurement); therefore, it can be used to directly compare dispersion across different data sets.
- **Interpretation of CV:** The greater the value of CV, the higher the risk.
- An inverse CV = $\left(\frac{\bar{X}}{S}\right)$ ➔It indicates unit of mean value (e.g. % of return) per unit of S.D.

Practice: Example 14, Volume 1, Reading 8.

7.8 The Sharpe Ratio

The Sharpe ratio for a portfolio **p**, based on historical returns is:

$$Sharpe\ ratio = \frac{Mean\ portfolio\ return - Mean\ Risk\ free\ return}{S.D.of\ Portfolio\ return}$$

$$S_h = \frac{\bar{R}_p - \bar{R}_F}{S_P}$$

- Excess return on Portfolio = Mean portfolio return – Mean Risk free return ➔ it reflects the extra return required by investors to assume additional risk.
- The larger the Sharpe ratio, the better the risk-adjusted portfolio performance.
- When Sharpe ratio is positive, it decreases with an increase in risk, all else equal.
- When Sharpe ratio is negative, it increases with an increase in risk; thus, in case of negative Sharpe ratio, larger Sharpe ratio cannot be interpreted as better risk-adjusted performance.
- When two portfolios have same S.Ds, then the portfolio with the negative Sharpe ratio closer to 0 is superior to other portfolio.
- However, when two portfolios have different S.Ds, then the portfolio with the negative Sharpe ratio closer to 0 ***cannot*** be interpreted as superior to other portfolio.

Ex-ante Sharpe Ratio: It is the forward-looking sharp ratio for a portfolio based on *expected* mean return, the risk-free return and the S.D. of return.

Limitation of Sharpe Ratio: It uses standard deviation as a measure of risk; however, Standard deviation is appropriate to use as a risk measure for symmetric distributions. Thus, it overstates risk-adjusted performance.

Practice: Example 15, Volume 1, Reading 8.

8. SYMMETRY AND SKEWNESS IN RETURN DISTRIBUTIONS

Symmetrical return distribution or Normal distribution: It is a return distribution that is symmetrical about its mean i.e. equal loss and gain intervals have same frequencies. It is referred to as normal distribution.

- A symmetrical distribution has skewness = 0

Characteristics of the normal distribution:

1) In a normal distribution, mean = median.
2) A normal distribution is completely described by two parameters i.e. its mean and variance.
3) Approximately:

- 68% of the observations lie between ± one standard deviation from the mean.
- 95% of the observations lie between ± two standard deviations.
- 99% of the observations lie between ± three standard deviations.

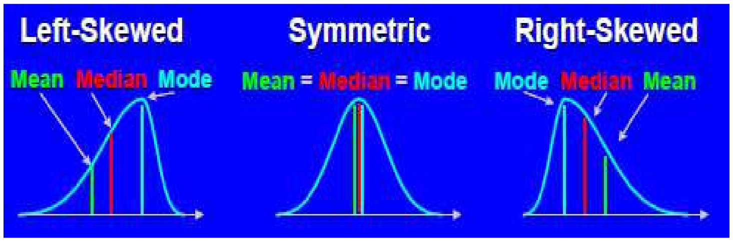
Skewed distribution: The distribution that is not symmetrical around the mean is called skewed.

a) Positively skewed or right-skewed Distribution: It is a return distribution that reflects *frequent* small losses and a *few* extreme gains i.e. ***limited*** but frequent downside.

- It has a long tail on its right side.
- It has skewness > 0.
- In a positively skewed unimodal distribution➔ mode < median < mean.
- Generally, investors prefer positive skewness (all else equal).

b) Negatively skewed or left-skewed Distribution: It is a return distribution that reflects *frequent* small gains and a *few* extreme losses i.e. ***unlimited*** but less frequent upside.

- It has a long tail on its left side.
- It has skewness < 0.
- In a negatively skewed unimodal distribution ➔ mean < median < mode.



Sample skewness (or sample relative skewness) is computed as follows:

$$S_K = \left[\frac{n}{(n-1)(n-2)}\right]\frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{S^3}$$

where,

n = *number of observations in the sample*
s = *sample S.D.*
n / (n-1)(n – 2) = *It is used to correct for downward bias in small samples.*

For larger values of n, sample skewness is computed as:

$$S_K \approx \left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{S^3}$$

- For n ≥ 100 ➔ a skewness coefficient of +/- 0.5 is considered unusually large.

Practice: Example 16, Volume 1, Reading 8.

9. KURTOSIS IN RETURN DISTRIBUTIONS

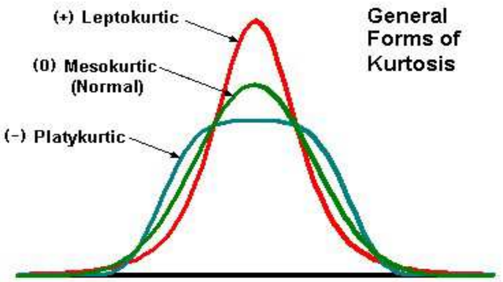
Kurtosis is used to identify how peaked or flat the distribution is relative to a normal distribution.

Leptokurtic: It is a distribution that is more peaked (i.e. greater number of observations closely clustered around the mean value) and has fatter tails (i.e. greater number of observations with large deviations from the mean value) than the normal distribution.

- It has more frequent *extremely* large deviations from the mean than a normal distribution.
- Ignoring fatter tails in analysis results in underestimation of the probability of extreme outcomes.
- The more leptokurtic the distribution is, the higher the risk.

Platykurtic: It is a distribution that is less peaked than normal.

Mesokurtic: It is a distribution that is identical to the normal distribution.



The Sample excess kurtosis is computed as:

$$K_E = \left(\frac{n(n+1)}{(n-1)(n-2)(n-3)}\frac{\sum_{i=1}^{N}(X_i - \bar{X})^4}{S^4}\right) - \frac{3(n-1)^2}{(n-2)(n-3)}$$

- For a normal distribution (mesokurtic), kurtosis = 3.0.
- For a leptokurtic distribution, kurtosis> 3.
- For a platykurtic distribution, kurtosis < 3.

NOTE:

Kurtosis is free of scale (i.e. it has no units of measurement).

It is always positive number because the deviations are raised to the 4th power.

Excess kurtosis = Kurtosis – 3

- A normal or mesokurtic distribution has excess kurtosis = 0.
- A leptokurtic distribution has excess kurtosis > 0.
- A platykurtic distribution has excess kurtosis < 0.

For larger sample size(n), Excess Kurtosis is computed using the following formula:

$$\frac{n^2}{n^3}\frac{\sum(X-\bar{X})^4}{S^4} - \frac{3n^2}{n^2} = \frac{1}{n}\frac{\sum(X-\bar{X})^4}{S^4} - 3$$

- For n ≥ 100 (taken from a normal distribution), a sample excess kurtosis of ≥ 1.0 would be considered unusually large.

Practice: Example 17, Volume 1, Reading 8.

10. USING GEOMETRIC AND ARITHMETIC MEANS

- For estimating single-period average return, arithmetic mean should be used.
- In contrast, for estimating average returns for more than one period, geometric mean should be used.

$$\textbf{Geometric mean return} \approx \textbf{Arithmetic mean return} - \frac{\textbf{\textit{Variance of return}}}{\textbf{2}}$$

Important to Note:

To plot past performance on a graph, it is more appropriate to use semi-logarithm scale rather than using arithmetic scale.

Semi-logarithm graph: In this graph,

- There is an ***arithmetic scale*** on the horizontal axis for time.
- There is a ***logarithmic scale*** on the vertical axis for the value of the investment.
- The values plotted on the vertical axis are gaped according to the differences between their logarithms.
  - o Suppose, values of investment are \$1, \$10, \$100 and \$1,000. Each value are equally spaced on a logarithm scale because the difference in their logarithms is equal i.e. ln10 – ln1 = ln100 – ln10 = ln1000 – ln100 = 2.30.
- On the vertical axis, equal changes between values represent equal % changes.
- The growth at a constant compound rate is plotted as a straight line i.e. upward (downward) sloping curve reflects increasing (decreasing) growth rates over time.

Important to Note:

- The arithmetic mean is appropriate to use for analyzing *future (or expected)* performance.
- In contrast, the geometric mean is appropriate to use for analyzing *past* performance.

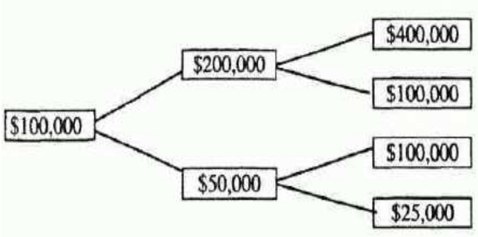
Example:

Suppose,

- Total amount invested = \$100,000
- Probability of earning 100% return = 50%.
- Probability of earning -50% return = 50%.
  - o With 100% return, return in one period = 100% × \$100,000 = \$200,000.
  - o With –50% return in the other period, return = –50% × \$100,000 = \$50,000

Geometric mean return =$\sqrt{(1+100\%)\times(1-50\%)}$ **–1 = 0**

With 50/50 chances of 100% or –50% returns, consider four equally likely outcomes i.e. \$400,000, \$100,000, \$100,000, and \$25,000.



Arithmetic mean ending wealth=(\$400,000 + \$100,000 + \$100,000 + \$25,000) / 4
= \$156,250.

- Actual returns are calculated as follows:
  - o $\frac{\$400{,}000-\$100{,}000}{\$100{,}000}\times 100 = 300\%$
  - o $\frac{\$100{,}000-\$100{,}000}{\$100{,}000}\times 100 = 0\%$
  - o $\frac{\$100{,}000-\$100{,}000}{\$100{,}000}\times 100 = 0\%$
  - o $\frac{\$25{,}000-\$100{,}000}{\$100{,}000}\times 100 = -75\%$

Arithmetic mean return for two-period = (300% + 0% + 0% – 75%) / 4
= 56.25%.

Arithmetic mean return for single-period = ((1+56.25 %)$^{1/2}$ –1)× 100 = 25%
≈ 25%

- According to this arithmetic mean return, arithmetic mean ending wealth = \$100,000 × 1.5625 = \$156,250.

Conclusion: *In order to reflect the uncertainty in the cash flows, the expected terminal wealth of \$156,250 should be discounted at 25% arithmetic mean rate not the geometric mean rate.*

Source: "10 Using Arithmetic and Geometric Means"
Volume 1, Reading 8.

Practice: End of Chapter Practice Problems for Reading 8.