A Statistical Refresher

1. A probability distribution lists the values of returns, may take the probabilities associated with different possible returns.

2. The expected value of a future return is also called the mean of the probability distribution and is calculated as the probability weighted average of the possible future values of that return. This is the sum of the probabilities of the events times the values of the return if those events occur.

3. The variance of the distribution of a future return is the probability weighted average of the squared deviations from the mean return.

4. A related concept is standard deviation, which is measured in the same units as returns because it is the square root of variance. The covariance of two return describes how the possible realizations of these returns move together.

   It is measured by the probability weighted average of the deviations from the mean for one return time the deviations from the mean of another return.

5. A related concept is the correlation coefficient the covariance of two returns, divided by the product of the two respective standard deviations.

   Correlations range between -1 and +1 .If two returns are perfectly correlated, the correlation is +1 if they are not at all correlated, the correlation is 0 and if they are perfectly negatively correlated, the correlation is -1

(Source for the Statistical Refresher: Mastering Investments, Part Three, 13th September, 2002, Special supplement with The Economic Times of India, Delhi)

1. MEASURE OF CENTRAL TENDENCY:
   1.1 Mathematical averages are
      a) Arithmetic mean (AM)
      b) Geometric mean (GM)
      c) Harmonic mean. (HM)

      The measure of central tendency called Median and Mode are Averages of Position.

   1.2 Arithmetic mean:
      1.2.1 For ungrouped data:
         a) Sigma f x divided by sigma f n. Here, f is frequency, 'x' is central value of class interval and n is total numbers.
         b) Also, A + sigma d divided by n. d = x minus A.
            Here A is the assumed mean.

      1.2.2 For discrete and continuous series:
         a) Sigma f x over sigma f.
         b) Also A + sigma f d divided by sigma f.
            Here d = x minus A.

2. MODE is the value of variate which occurs most frequently in an ungrouped data. For continuous series: Mode = L + (f_0 minus f_0 divided by 2 f_1 minus f_0 minus f_2) into small i where L is the lower limit of class interval in which mode is situated; f_1 is the maximum frequency; f_0 is frequency of the preceding class interval, f_2 is the frequency of succeeding class interval ; i is the length of the class interval.

   When distribution is not normal: Mode = 3 median minus 2 mean (mean minus mode = 3 (mean – median)

   AM > GM > HM

   $(AM)(HM) = (GM)^2$

3. MEDIAN: which is also the positional average is the middle term of the array, when the ungrouped data is arranged in ascending or descending order.
   a) For ungrouped data:

Median is size (u+1/2) item if n is an odd number. If n is an even number, then median = [size of n/2 item + f (n/2 +1) item] divided by 2.

b) For continuous series

Median = L+ small i / f (n/2 minus c) where

n = total frequency;

L = lower limit of grade in which median number lies;

i = class interval;

f = frequency of the median group and

c = the cumulative frequency of the group preceding the median group.

4. Partition values: If the division of a series is into
   a) 4 equal parts, each part is called Quartiles. The first Q1 = n/4; Q3 is 3n/4. For grouped data: Q1 = l + small I /f (n/4 minus c).
   b) 10 equal parts, each part is called deciles;
   c) 100 equal parts, each part is called percentiles.

5. For Geometrical Mean: average rate of increase of variate.
   a) Ungrouped data: GM = (x_1 minus x_2 into x_3 ….. $x_n)^{1/n}$ of GM = i/n sigma 4x. Here x_c is the mid value of the class interval.
   b) Discrete series and continuous series, GM = sigma f lgx / sigma f where f = corrected frequency of the variable x of variate.
6. Harmonic Mean: average rate of increase of variate.
   a) Ungrouped data: HM = n divided by $1/x_1 + 1/x_2 + \ldots + 1/x_n$
   b) Discrete series and continuous series: HM = sigma f / sigma (f/x)