Lesson 2: Measures of Central Tendency, Other Measures of Location (Quantiles) and Measures of Dispersion

Measures of Central Tendency

- Look to identify the “middle” or “expected value” of a data set.

Arithmetic Mean

- Most frequently used measure of central tendency.
- Equals the sum of all the observations in a data set divided by the total number of observations.
- Can be calculated for the entire population and for a sample.

Example

The scores of 20 students on a 100-point exam are given as 77, 90, 57, 85, 68, 31, 45, 86, 46, 98, 25, 10, 57, 67, 88, 77, 34, 89, 47, and 77. Calculate the population mean and the mean for a sample including only the first 5 observations.

Properties of the Arithmetic Mean

- All observations are used in the computation of the arithmetic mean.
- All interval and ratio data sets have an arithmetic mean.
- The sum of the deviations from the arithmetic mean is always 0.
- An arithmetic mean is unique i.e., a data set only has one arithmetic mean.

Problem with the Arithmetic Mean

- Very sensitive to extreme values.

Median

- The value of the middle item of a data set once it has been arranged in ascending or descending order.
  - For a data set where the total number of observation is an odd number, the median is the value of the item in [(n+1)/2] position.
  - In an even-numbered data set, the median is the average of the observations occupying positions (n/2) and [(n+2)/2].

Advantage

- It is not sensitive to extreme values.

Disadvantage

- It does not use all information about the size and magnitude of observations and only focuses on their relative positions.

Mode

Equals the most frequently occurring value in a data set:

- A data set that has one mode is said to be unimodal.
- A data set that has two modes is said to be bimodal.
- It is also possible for a data set to have no mode.
- For grouped data, the modal interval is the interval with the highest frequency.

The Weighted Mean

- Calculated by assigning *different* weights to observations in the data set to account for the disproportionate effect of certain observations on the arithmetic mean.
- The arithmetic mean assigns an *equal* weight to every observation in the data set, which makes it very sensitive to extreme values.

Example

An individual invests 30% of her portfolio in Stock A, 40% in Stock B and 30% in Stock C. The expected returns on Stock A, B and C are 10%, 14% and 6% respectively. What is the portfolio's expected return?

The Geometric Mean

- Used to average rates of change over time or to calculate the growth rate of a variable over a period.

- To calculate the geometric mean for investment returns data, add 1 to each return observation (expressed as a decimal) and subtract 1 from the result.

Example

The returns on XYZ common stock in the years 2000, 2001, 2002 and 2003 were 15.3%, 6.7%, -10% and -2.3% respectively. Compute the geometric mean of these returns.

Important Relationships Between the Arithmetic Mean and the Geometric Mean

- The geometric mean is always less than, or equal to the arithmetic mean.
- The geometric mean equals the arithmetic mean only when all the observations are identical.
- The difference between the geometric and arithmetic mean *increases* as the dispersion in observed values increases.

The Harmonic Mean

- Used in the investment management arena to determine the average cost of shares purchased over time.
- Can be viewed as a special type of weighted mean, where the weight of an observation is inversely proportional to its magnitude.

Example

An investor purchased $5,000 worth of RPS Stock each month over the last 4 months at prices of $4, $5, $6 and $7. Determine the average cost of the shares acquired.

A value at, or below which a stated proportion of the observations in a data set lie.

- Quartiles divide the distribution in quarters or four equal parts.
- Quintiles divide the distribution into fifths.
- Deciles divide the data into tenths.
- Percentiles divide the distribution into hundredths.

Example

1. Calculate the first quartile of a distribution that consists of the following asset returns: 10%, 23%, 13%, 17%, 19%, 5%, 4%.
2. If we include one more return observation of 10% in our data set, what is the new value of the first quartile?

MEASURES OF DISPERSION

Dispersion refers to the variability or spread of a random variable around its central tendency.

- The mean represents the expected reward (return) from an investment.
- The dispersion measures the risk in the investment.

Range

- The difference between the highest and lowest values in a data set.

Example

Calculate the range of the following data set that contains average returns earned by a portfolio manager over the last 6 years: 4%, 1%, 8%, 10%, 3%, 15%.

Mean Absolute Deviation (MAD)

- Average of the *absolute* values of deviations of observations in a data set from its mean.

Example

Calculate the MAD of the following data set that contains average returns earned by a portfolio manager over the last 6 years: 4%, 1%, 8%, 10%, 3%, 15%.

Population Variance and Standard Deviation

The **variance** equals the sum of the squares of deviations from the mean.

- Has no unit.

The **standard deviation** is the positive square root of the variance.

- Expressed in the same units as the random variable itself.

The variance calculated using all the observations in a population is called the **population variance**.

The **population standard deviation** is the positive square root of the population variance.

Example

Calculate the variance and standard deviation of the scores of five golfers assuming that they represent the entire population of golfers participating in a particular tournament. Their scores are 67, 71, 72, 75 and 68.

Sample Mean

Sample Variance

Sample Standard Deviation

Example

Calculate the variance and standard deviation for the golfers' scores in Example 2-9. However, assume that the 5 scores that are presented are the scores of a representative sample of golfers from the entire population of golfers who participated in the tournament.

When only concerned with *downside risk* (e.g. the risk of obtaining returns below the mean) it is quite helpful to compute:

- **Semi variance**, which is the average of squared deviations below the mean.
- **Semi deviation**, which equals the positive square root of the semi variance.

Chebyshev's Inequality

- Calculates an approximate value for the proportion of observations in a data set that lie within *k* standard deviations from the mean.

Example

Determine the minimum percentage of observations in a data set that lie within 1.5 standard deviations from the mean.

Advantage

It holds for samples and populations and for discrete and continuous data regardless of the shape of the distribution.

The standard deviation is more easily interpreted than the variance because it is expressed in the same units as the data.

However, it may sometimes be difficult to determine what standard deviation means in terms of relative dispersion of the data.

- The data sets being compared have significantly different means.
- When the data sets have different units of measurement.

In such situations, you might want to use measures of **relative dispersion**.

Coefficient of Variation

- Ratio of the standard deviation of the data set to its mean.
- Used to measure the *risk per unit of return* in various investments.

Example

Calculate the coefficient of variation for the following stocks and interpret the results:

	Mean Return	Standard Deviation
Stock A	12%	13%
Stock B	5%	4%
Stock C	10%	8%

The Sharpe Ratio

- Ratio of excess return over the risk-free rate from an investment to its standard deviation of returns.
- Basically measures *excess return per unit of risk.*
- Portfolios with *higher, positive* Sharpe ratios are *more* attractive as they offer higher excess returns per unit of risk.

Example

Calculate the Sharpe ratio for a portfolio with a mean return of 12.5% and standard deviation of 10% given that the risk-free rate is 6%.