Summarization of Data by Moments

Moments
Used to describe several properties of a distribution

Moment about zero

\[ \frac{1}{N} \sum_{i=1}^{N} X_i^r \]

Moment about mean

\[ \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^r \] where the exponent "r" defines the r\text{th} moment

Various moments
• First moment about zero – Mean
• Second moment about zero – Variance
• Third moment about zero – Skewness
• Fourth moment about mean – Kurtosis
1st Moment of Data - Mean

Mean

- The mean of data is equal to the sum of n observations divided by the number of observations
- Mean has a property that sum of the deviation from each score is always zero

Mathematically

\[ \text{Mean} = \frac{\sum_{i=1}^{n} X_i}{n} \]

By convention

- \( \bar{X} \) is used to denote the sample mean
- \( \mu \) is known as population mean

X percent trimmed mean

- Is the mean is the mean after removing x percent of the data from the upper and lower extremes
- The 5% trimmed mean is calculated after 5% of the observations from the upper and lower extremes has been removed
- Use in order to remove the effect of extreme observations
2nd Moment of Data - Variance

Variance
• the sum of each observation minus the mean, squared, divided by the sample size
• Measures the average squared deviation from the mean
• Quantity to measure dispersion, how much a particular variable deviates

Mathematically, $\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$

In case the sample size is very small as compared to the population, we use n-1 rather than n, in that case

$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n - 1}$

Standard deviation
• Positive root of is known as standard deviation of the data
• The standard deviation is the average deviation from the mean, expressed in the original units.
• Mathematically, $\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$
Properties of Variance

Let $X$ and $Y$ be two different random variables and $c$ any real number. Then,

- $\text{Var } c = 0$ (i.e. variance of any constant number is zero)
- $\text{Var } (c+X) = \text{Var}(X)$
- $\text{Var } (X+Y) = \text{Var } (X) + \text{Var } (Y)$
- $\text{Var } (X-Y) = \text{Var } (X) - \text{Var } (Y)$
- $\text{Var}(X+c) = \text{Var}(X)$
- $\text{Var}(aX+c) = a^2 \text{Var}(X)$
- $\text{Var}(aX+bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Covariance}(X,Y)$
Negative Semi Variance

- Similar to variance but only negative deviations from mean are used in the calculation.

- Logic behind the calculation of Negative Semi-Variance is that the positive deviations doesn't constitute the actual risk for the investor. Only risk for the investor is when the returns go below the benchmark rate. So, it basically measures the downside risk, which actually matters,

\[ \text{NegativeSemi - Variance} (\sigma^2) = \frac{\sum (x_i - \mu)^2}{m - 1} \]

Where, 
- \( m \) is the total number of returns lower than the mean.
- \( \mu \) is the mean of the observations.
Higher Order Moments – Skewness (3rd Moment)

• Measures the degree of asymmetry exhibited by the data

Mathematically,

Measure of Skewness \( k = \frac{\sum (x - \bar{x})^3}{n - 1} \left( \frac{\sum (x - \bar{x})^2}{n - 1} \right)^{\frac{3}{2}} = \frac{\sum (x_i - \bar{x})^3}{n \sigma^3} \)

• If skewness is equal to zero, the histogram is symmetric about the mean
  – Positive and negative skewness

**Negative-Skewed**

Mean, Median, Mode

**Symmetric**

Mean = Median = Mode

**Positive-Skewed**

Mode, Median, Mean
...Higher Order Moments – Skewness & Kurtosis

- Kurtosis describes the degree of "flatness" of a distribution, or width of its tails. High kurtosis indicates a higher probability of extreme movements.
- Mesokurtotic
  - kurtosis is equal to three
  - normal distribution has this kurtosis
- Platykurtic
  - kurtosis < 3
  - the curve is more flat and wide, called negative kurtosis
- Leptokurtic
  - kurtosis > 3
  - the curve is more peaked, called positive kurtosis
  - often found in asset returns when there are periodic jumps in asset prices, which happens frequently in markets where there is discontinuous trading such as security markets that close overnight or at weekends

\[ k = \frac{\sum (x - \bar{x})^4}{n - 1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{n \sigma^4} \]