SFM-02 Summary Document

Probability:

Calculating Probability: Probabilities for events are obtained by counting the occurrence of events and total observations.

Probability of an event happening = Number of outcomes for which event is True / Total number of outcomes

Calculating Probability for a group of Events:
Pairs (or groups) of events: A and B
One or the other or both occur: A or B ≡ A ∪B
Both events occur A and B ≡ A ∩ B

P(A ∪B) = P(A) + P(B) - P(A ∩ B)

Example: Rolling a **couple** of dice:

A Probability that even number shows up {2,4,6}
B Probability that odd number shows up {1,3,5}
p(A) = p(occurence of 2) + p(occurence of 4) + p(occurence of 6) = ⅙+⅙+⅙ = ½
p(B) = p(occurence of 1) + p(occurence of 3) + p(occurence of 5) = ⅙+⅙+⅙ = ½
p(A and B) = ½ * ½ = ¼
p(A U B) = p(A) + p(B) - p(A ∩ B)
p(A U B) = ½ + ½ - ¼

Independent events: Occurrence of A does not affect the probability of B.
p(A ∩ B) = P(A) * P(B)

Mutually Exclusive events: Either event A or event B can occur, but they can not occur together
P(A ∩ B) = 0

Summary of Prob Events:

Category of Event	Type of Event	Calculation

Single	Event A	P(A)
Single	non-Event A	1 - P(A)
Multiple	Event A or B(Mutually Exclusive)	P(A) + P(B)
Multiple	Event A or B(non-Mutually Exclusive)	P(A) + P(B) - P(A ∩ B)
Multiple	Event A and B	P(A ∩ B)
Multiple	Event A given occurrence of Event B	P(A|B) = P(A ∩ B) / P(B)

Probability Distribution and Density:

Probability Distribution: is a set of probabilities of events that can arise from a series of observations.

Probability mass function:
∑P(x)=1
x=(all possible observations)

Sample and Population :
A sample dataset is a subset of the population dataset. We generally refer to the sample dataset for most calculations.

Expected Value:
The expected value of a discrete random variable is the weighted mean of the variable's all possible values. The weight is the probability of the random variable for a specific value.
Expected Value = Sum of(Possible Values * respective probabilities)

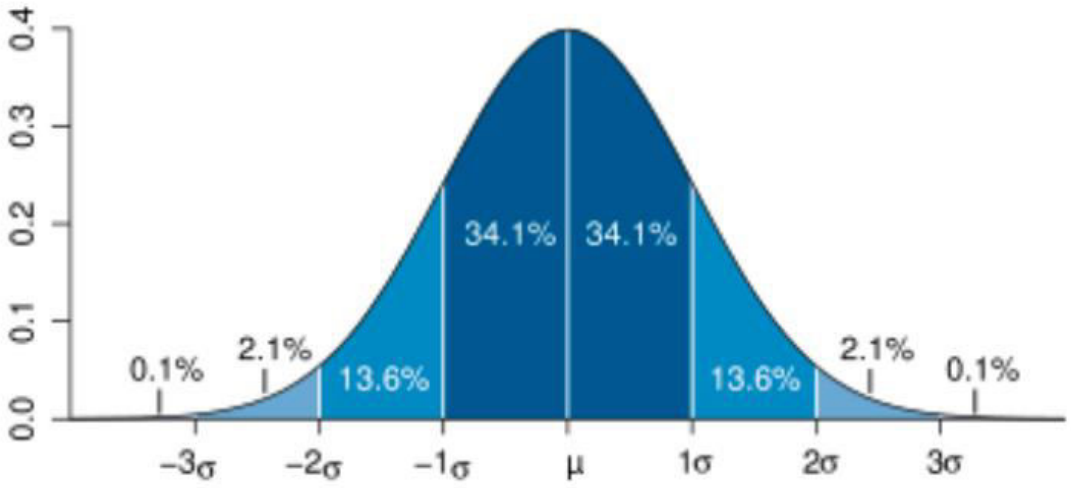
What is the expected value when you roll a dice once:
EV = 1*(1\6) + 2*(1\6) + 3*(1\6) + 4*(1\6) + 5*(1\6) + 1*(1\6) = 3.5

The probability distribution is the list of all possible outcomes and their probabilities for a given event.

The probability distribution of a continuous random variable is called the probability density function. A probability for the occurrence of any number of a continuous variable is zero. Because there could be infinite numbers of possible outcomes in a continuous variable. But the probability density could be more than zero.

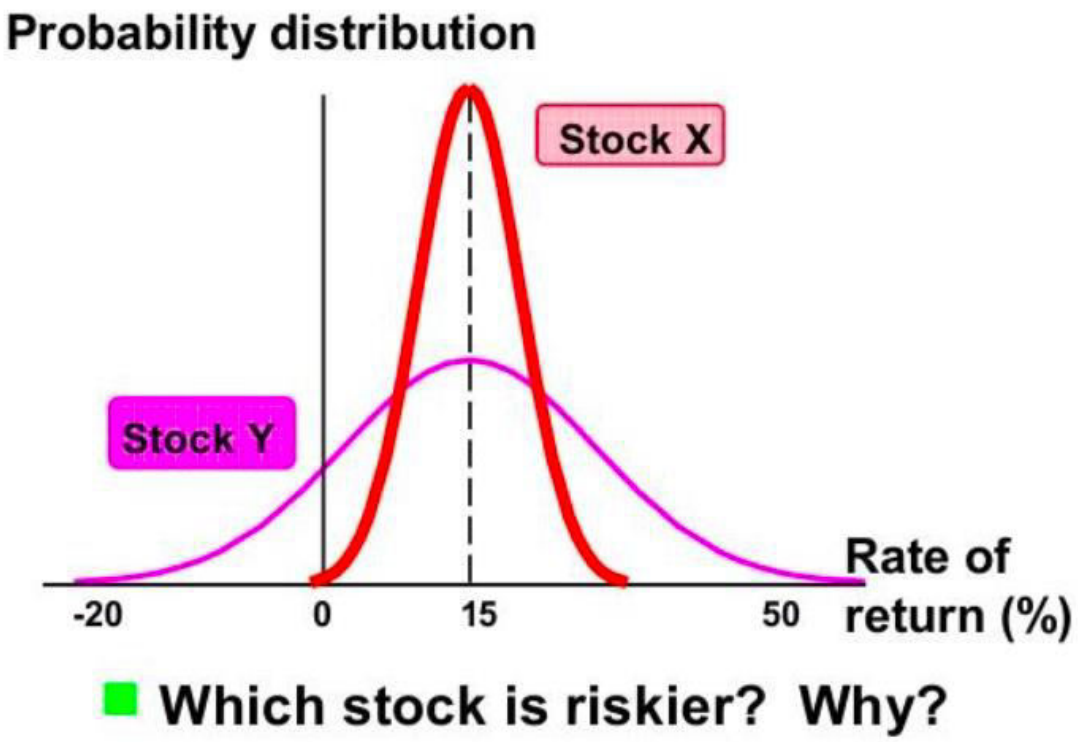
One of the most important probability distributions is Normal distribution, it follows few characteristics:

A. The area covered under standard deviation -1 and 1 attributes to 68% of the total area of the distribution.
B. The area covered under standard deviation -2 and 2 attributes to 95% of the total area of the distribution.
C. The area covered under standard deviation -3 and 3 attributes to 99% of the total area of the distribution.



Standard deviation is also considered as the volatility of the instrument and the measure of the risk in investing in that instrument.

Higher the standard deviation implies a higher risk involved. Similarly Lower the standard deviation implies a lower risk involved.



Basic Statistics

Mean: The mean is the average number of the dataset.

Median: The median separates the higher half and lower half of the dataset into an equal number of elements. The closest element between the higher half and lower half is the median.

Standard deviation: Standard deviation is the measure of the spread of the data from the mean. It is the degree of scatter of data around the mean.

Variance: Variance is the square of the standard deviation.

Capital Asset Pricing Model: CAPM is a way to measure systematic risk. It displays the relationship of the return of an asset with systematic risk. Systematic risk includes inflation, economic slowdown, recession, interest rate etc.

Linear regression: Linear regression is the way to develop a model by finding out the relationship between the two or more variables in a linear equation.

R-squared: R-squared is the statistical measurement of the fitment of the data points with the regression line.

The beta of the stock: Beta is the stock's volatility compared to the overall market. The benchmark index of the market generally compares it. For example, if we compute the beta of the S & P 500 index. It will come out to be 1 because we are computing the volatility of the benchmark index(S&P 500) with the volatility of the asset S&P 500 itself.

Hypothesis Testing

Hypothesis Testing: Hypothesis Testing is the way of testing an assumption using a sample dataset. Hypothesis testing checks the probability of a hypothesis considering the NULL hypothesis(inverse of the hypothesis) is True.

To test an assumption, let us define two statements :
Ha(Alternate Hypothesis): The assumption which we hope to be True.
H0:(Null Hypothesis): It is a mutually exclusive statement or assumption from the Alternate Hypothesis.

p-value: p-value is the probability of observing a more extreme value of the test statistic(than the one observed), provided the null hypothesis is true.

Significance level: Significance level is used to compare the p-value to accept or reject the NULL hypothesis.