Basic Statistics

Raison d'être

- Discovering Relationships *– Regression, Co-Integration, and other statistical relationships.*
- Measuring the Financial world *– mean returns, variance of returns, understanding risk, and expected return.*
- Beating the Odds *– Market as a casino. Discovering your edge is what it's about. Sizing it up – The Kelly Fraction.*
- Playing Games *– Arbitrage or Directional bet.*
- Thinking Smart *- better than an average Joe. Quantitative understanding, Patterns, or Hunches.*
- Getting Rich(er) *- Hopefully*

Basic Statistics

Probability

Decision Making Under Uncertainty

- Understanding probability.
- Using probability to understand expected value and risk.
- Applications
  - –Financial transactions at future dates.

… Life is full of uncertainty (*and the night is dark and full of terrors*)

Counting Rule for Probabilities

- Probabilities for events are obtained by counting.
- P(Event) = (Count of Event = TRUE)/ (Total Number of Event Outcomes)

Joint Events (Primer)

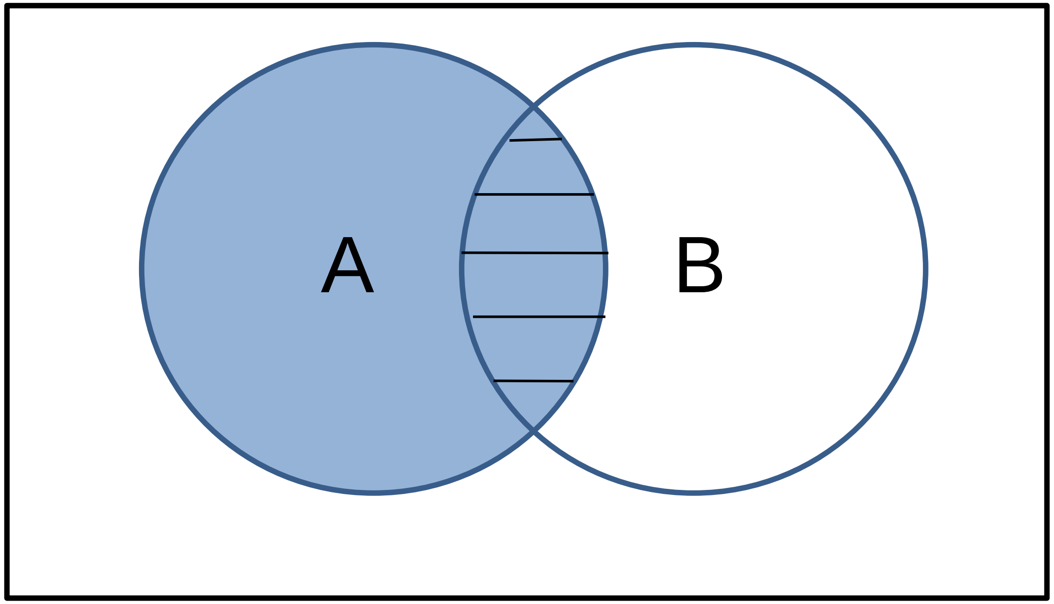
- Pairs (or groups) of events: A and B

  One or the other occurs: A **<u>or</u>** B $\equiv A \cup B$

  Both events occur A **<u>and</u>** B $\equiv A \cap B$

- An **addition rule**: $\boxed{P(A \cup B) = P(A)+P(B)-P(A \cap B)}$

Joint Events – Addition (Primer)



$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

Example – Roll a dice (Addition)

A → Probability that even number shows up {2,4,6}
B → Probability that '3' or '4' or '6' shows up {3,4, 6}

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$p(A \cup B) = 1/2 + 1/2 - 2/6 = 2/3$$

Independent Events

- **Independent events**: Occurrence of A does not affect the probability of B
- The **product rule** for independent events:

$$\boxed{P(A \cap B) = P(A)P(B)}$$

Example – Toss two coins (Multiplication)

A → Probability that 'Head' shows up in the first coin {H}

B → Probability that 'Tail' shows up in the second coin {T}

$$p(A \cap B) = p(A) * p(B)$$

$$p(A \cap B) = 1/2 * 1/2 = 1/4$$

Conditional Probability

- "Conditional event" = occurrence of an event given that some other event has occurred.
- Conditional probability = Probability of an event given that some other event is certain to occur. **Denoted P(A|B) = Probability of A will occur given B occurred.**
    - Prob(A|B) = Prob(A and B) / Prob(B)

Example – Bayes' Theorem

A → Probability that '2' shows up {2}

B → Probability that even number shows up {2,4,6}

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}$$

$p(A \cap B) \longrightarrow 1/6$

$p(B) \longrightarrow 3/6$

$= 1/3$

Example – Bayes' Theorem

A → Probability that '5' shows up {5}

B → Probability that even number shows up {2,4,6}

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}$$

$p(A \cap B) \longrightarrow 0 \quad = 0$

$p(B) \downarrow 1/2$

Using Conditional Probabilities: Bayes Theorem

$$P(A\,|\,B) = \frac{P(A,B)}{P(B)} \qquad \text{Target}$$

$$= \frac{P(B\,|\,A)P(A)}{P(B)} \qquad \text{Theorem}$$

$$= \frac{P(B\,|\,A)P(A)}{P(A,B) + P(notA,B)} \qquad \text{Definition}$$

$$= \frac{P(B\,|\,A)P(A)}{P(B\,|\,A)P(A) + P(B\,|\,notA)P(notA)} \qquad \text{Computation}$$

Example – Effect of Interest rates on Index

Interest Rate

- Increased 1000 days
  - Index Price
    - Increased 150 days
    - Decreased 850 days
- Decreased 1000 days
  - Index Price
    - Increased 700 days
    - Decreased 300 days

Total Sample days where Interest rate has been changed: 2000

Example – Effect of Interest rates on Index

Ques: Given that the Fed has increased the interest rate what is the probability that the Index will decrease?

A → Probability of Index price decreasing
B → Probability of interest rate increasing

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \longrightarrow \frac{850/2000}{1000/2000} = 85\%$$

Expected Value

Expected Value = SUM of {all values * their probability}

Q: What is the expected value when you roll a dice once?

$$EV = 1/6 * 1 + 1/6 * 2 + 1/6 * 3 + 1/6 * 4 + 1/6 * 5 + 1/6 * 6$$

$$EV = 3.5$$

Q: A stock can increase by 20% with a prob. of 70% and decrease by 20% with a prob. of 30%. What is the expected value of the stock currently trading at $100?

$$EV = 120 * 0.7 + 80 * 0.3$$

Fair Value of Game? (Expected Value)

A → If '5' or '6' shows up payout is \$100 (total profit is \$100)
B → If any other number shows up payout is zero

x → Fair value of the game

Winnings = 1/3 * \$100

Losses = 2/3 * x

Fair value, x = \$50

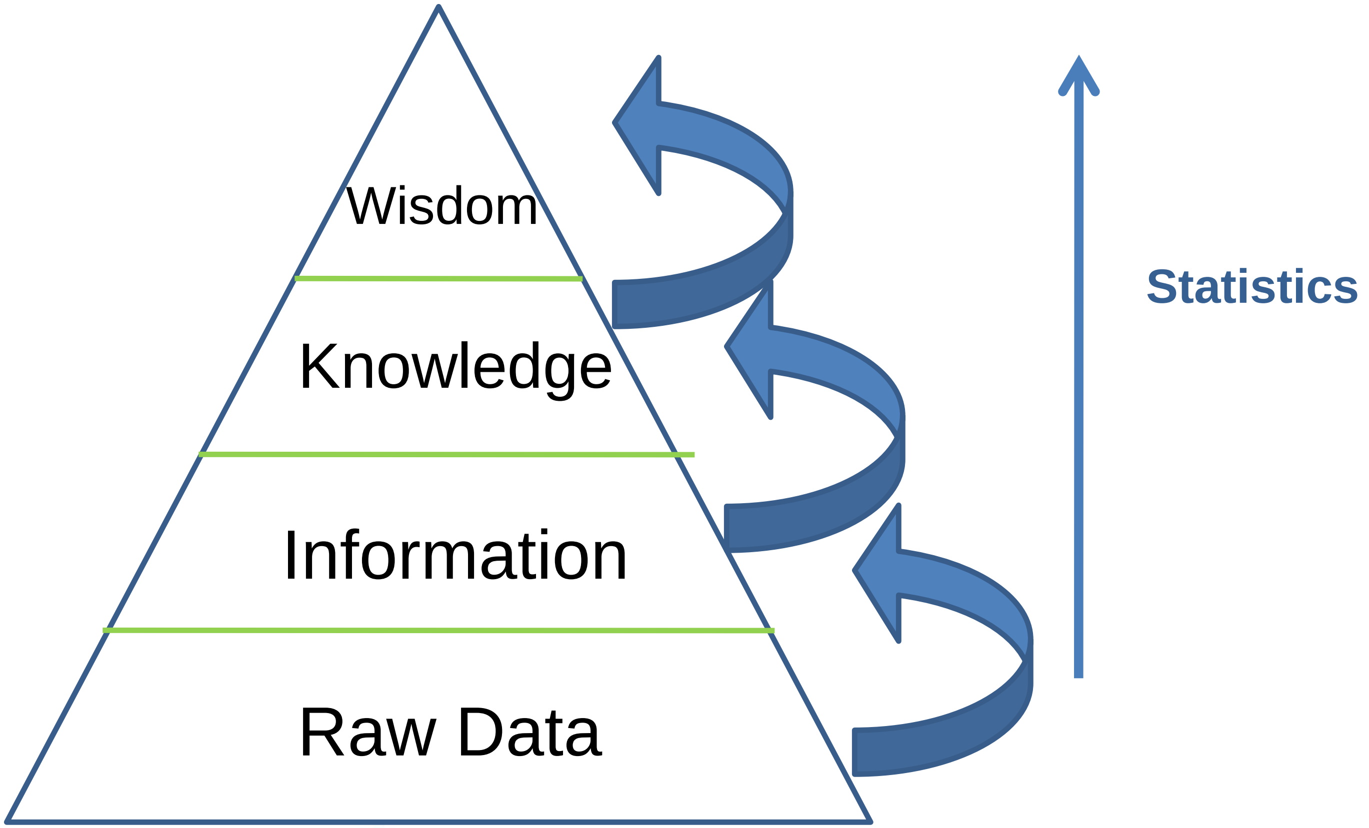
Q & A

Basic Statistics

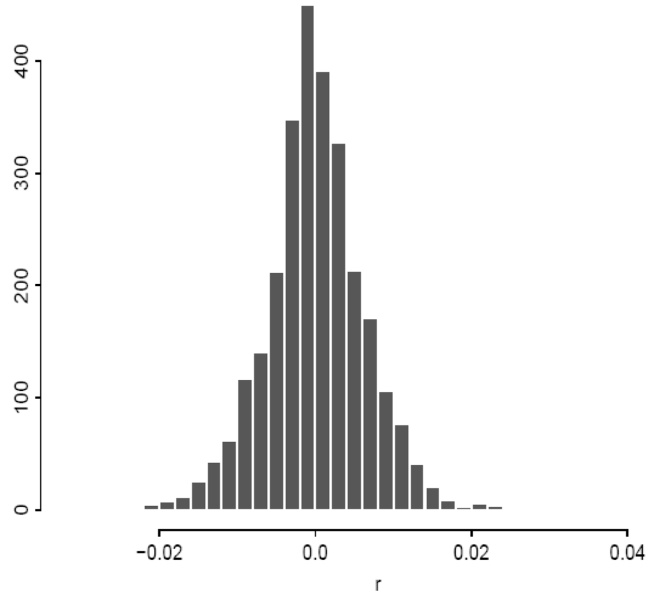
Visualizing the Data –*ask the right questions*

- **What story does the data presentation tell?**
  - **Data in raw form tell no story.**
  - **Visual representation of data may tell something about the data.**
- **Data reduction and summary representation: What do we learn?**
  - **Location (*Mean*).**
  - **Spread (*Variance*).**
  - **Shape of the distribution (*probability density function*).**
- **What tool is most informative?**
  - **Reduction to a small number of features (*Summary of data*).**
  - **Visual displays of data (*Plot*).**
    - **Histograms.**
    - **Time series plots – line charts.**
- **What hypothesis are most likely?**
  - **Mean Reversion?**
  - **Trending?**
  - **Any likely relationship between variables - guesses?**

Data Pyramid



Visualizing the Data –*examples*



Visualizing the Data – *non financial world*

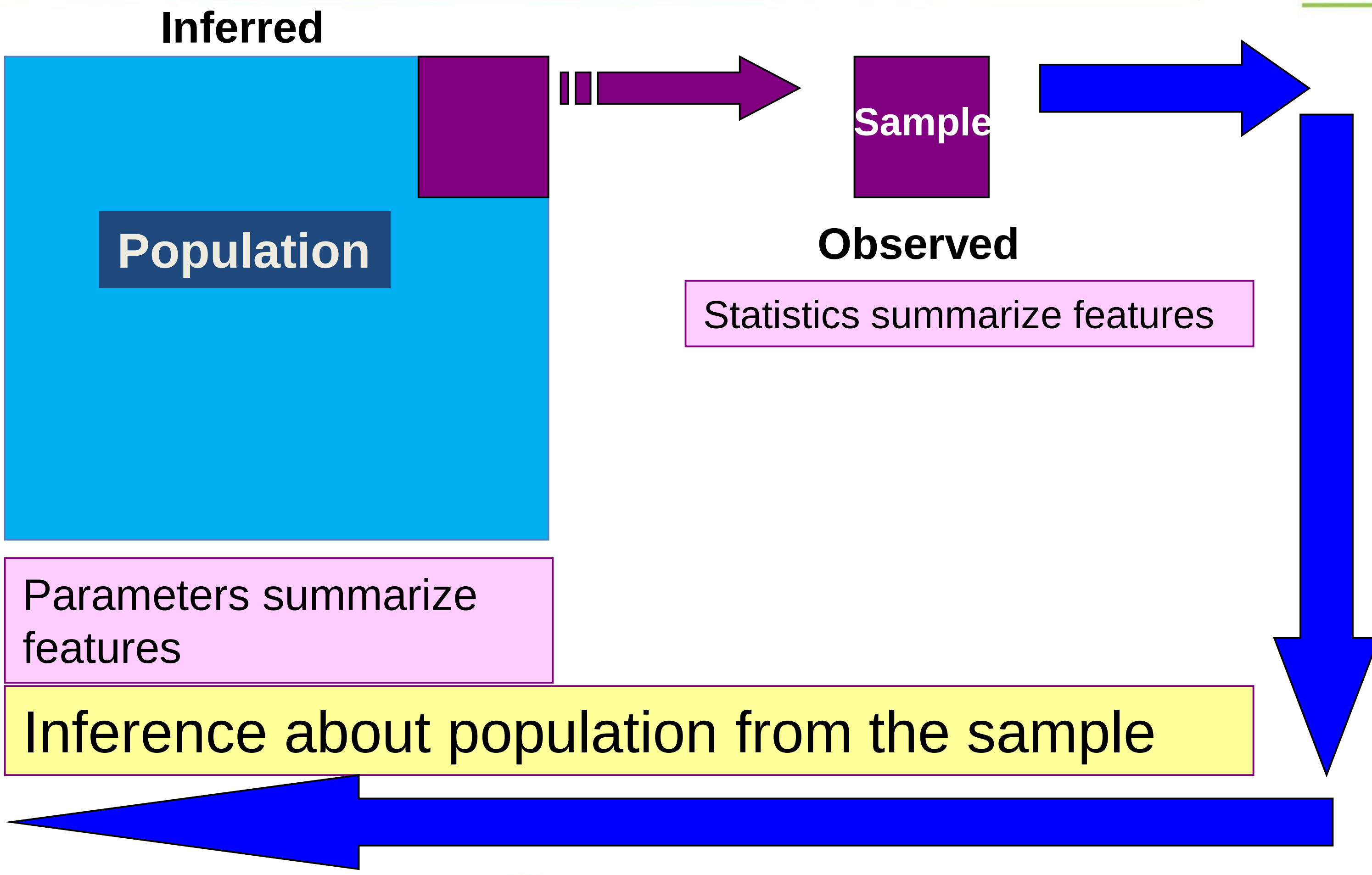
- ***http://blog.ted.com/2007/06/26/hans_roslings_j_1/***
- ***Google for "motion charts".***
- ***It is very unlikely that we will get a chance to create a motion chart in the present course.***

WARNING - "There are lies, damned lies and statistics." (Benjamin Disraeli)

Sample and Population

- What we have is a sample .
- The sample “looks like” the population.
- The bigger is the RANDOM sample, the closer will be the resemblance.
- The population is (effectively) infinite and the sample is a trivial proportion of the population. How can a population be infinite? The subject of the rest of this course.
- Sample has a sample mean and standard deviation.
- Population has a mean- μ, and standard deviation- σ.
- The sample statistics resemble the population features.

Sample and Population



Q & A

Basic Statistics

Moments of Data – Mean and Variance

Mean, Variance

- Mean is also known as
  - The average.
  - The arithmetic mean.
  - To calculate we take the sum of all values and divide it by the number of data values.

- The Median divides the distribution into two halves.
- Range is the difference between the largest and the smallest value in the data.
- Standard deviation is measure of spread of the data from the mean. It is the degree of scatter of data around the mean.
- Variance is the square of the standard deviation.
- All these statistics are used to summarize the data. from a sample.

In the Financial World

- Mean Returns – are measure of expected return.
- Variances – measure of risk.

Returns and Risk of a 2 stock portfolio

We have two stocks, X and Y, and we know the average annual returns and the annual standard deviation of these stocks.

If we allocate equal capital in both stocks what would be the returns and standard deviation of the portfolio?

Mean = E[X]

Mean of [X+Y] = E[X+Y] = E[X]+E[Y]

Mean of [aX + bY] = E[aX] + E[bY]
= aE[X] + bE[Y]

Mean, Variance of sum of variables.

We have two stocks, say SBIN and PNB, and we know the average annual returns and the annual standard deviation of these stocks.
If we allocate equal capital in both stocks what would be the returns and standard deviation of the portfolio?

$$\text{Variance} = \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$$

Variance[x+y] = Var[x] + Var[y] +2Cov(x,y)
Cov(x,y) = $\rho_{xy} * \sigma_x * \sigma_y$ where correlation coefficient is ρ_{xy}

Variance[ax+by] = (a^2)Var[x] +(b^2)Var[y] +2 * a * b * $\rho_{xy} * \sigma_x * \sigma_y$

Looking ahead for two variables – covariance, correlation

Variables Y and X vary together
Does movement in X "cause" movement in Y in some sense?
Covariance - Simultaneous movement through a **statistical relationship**

$$\text{Standard Deviations: } S_X = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}, \quad S_Y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}$$

$$\text{Covariance: } s_{XY} = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{N-1}$$

$$\text{Correlation: } r_{XY} = \frac{\sum_{i=1}^{N}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right)}{N-1}$$

$$= \frac{s_{XY}}{s_X s_Y}$$

$\boxed{-1 \le r_{XY} \le +1 \quad \textbf{Units free. A pure number.}}$

Basic Statistics

Application - Capital Asset Pricing Model

Portfolio Management Example

- You have \$1000 to allocate between assets A and B. The yearly returns on the two assets are random variables r_A and r_B.
- The means of the two returns are $E[r_A] = \mu_A$ and $E[r_B] = \mu_B$
- The standard deviations(risks) of the returns are σ_A and σ_B.
- The correlation of the two returns is ρ_{AB}

- You will allocate a proportion w of your \$1000 to A and (1-w) to B.

Version 10.0.1

Portfolio Management Example

- Suppose you know μ_A, μ_B, ρ_{AB}, σ_A, and σ_B (You have watched these stocks for many years.)
- The mean and standard deviation are then just functions of w.
- You will then compute the mean and standard deviation for different values of w.
- For example, $\mu_A = .04$, $\mu_B = .07$

  $\sigma_A = .02$, $\sigma_B = .06$, $\rho_{AB} = -.4$

$$E[\text{return}] = w(.04) + (1-w)(.07)$$
$$= .07 - .03w$$

$$SD[\text{return}] = \text{sqr}[w^2(.02^2) + (1-w)^2(.06^2) + 2w(1-w)(-.4)(.02)(.06)]$$
$$= \text{sqr}[.0004w^2 + .0036(1-w)^2 - .00096w(1-w)]$$

Return and Risk

- Your expected return is

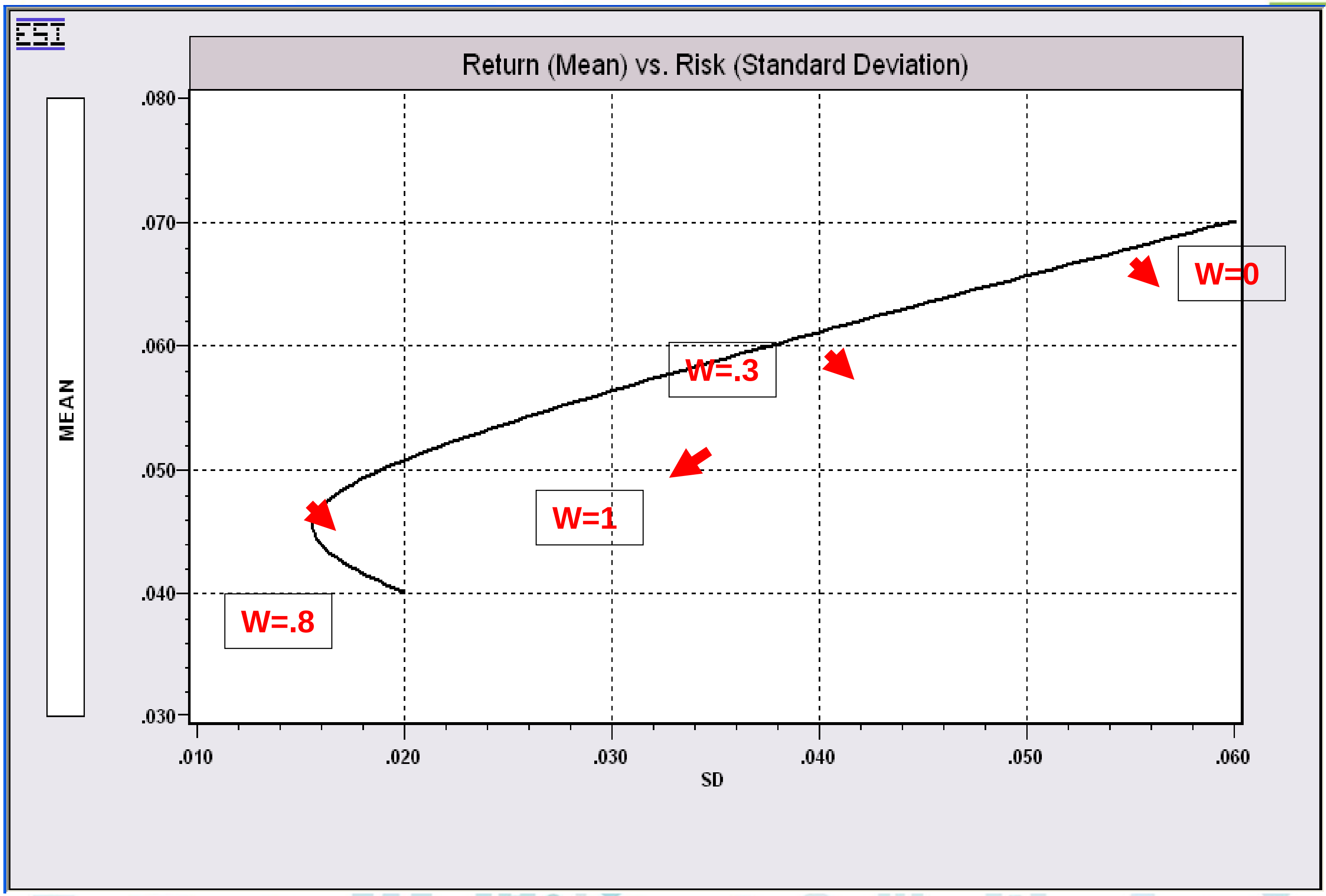
$$E[wr_A + (1\text{-}w)r_B] = w\mu_A + (1\text{-}w)\mu_B$$

- The variance of your return is

$$Var[wr_A + (1\text{-}w)r_B] = w^2\sigma_A^2 + (1\text{-}w)^2\sigma_B^2 + 2w(1\text{-}w)\rho_{AB}\sigma_A\sigma_B$$

- The standard deviation is the square root of variance.

Risk vs. Return



Probability Distribution Function

Summarizes how odds/probabilities are distributed among the events that can arise from a series of observations.

Probability density function for a discrete case (P(dice shows 3))

It would be best to use a table to represent the probability distribution

Typically called probability mass function

Probability density function for a continuous case (P(0.03>ret>0.01))

It would be best to use a function or a chart to represent the probability distribution. Typically called the probability density function.

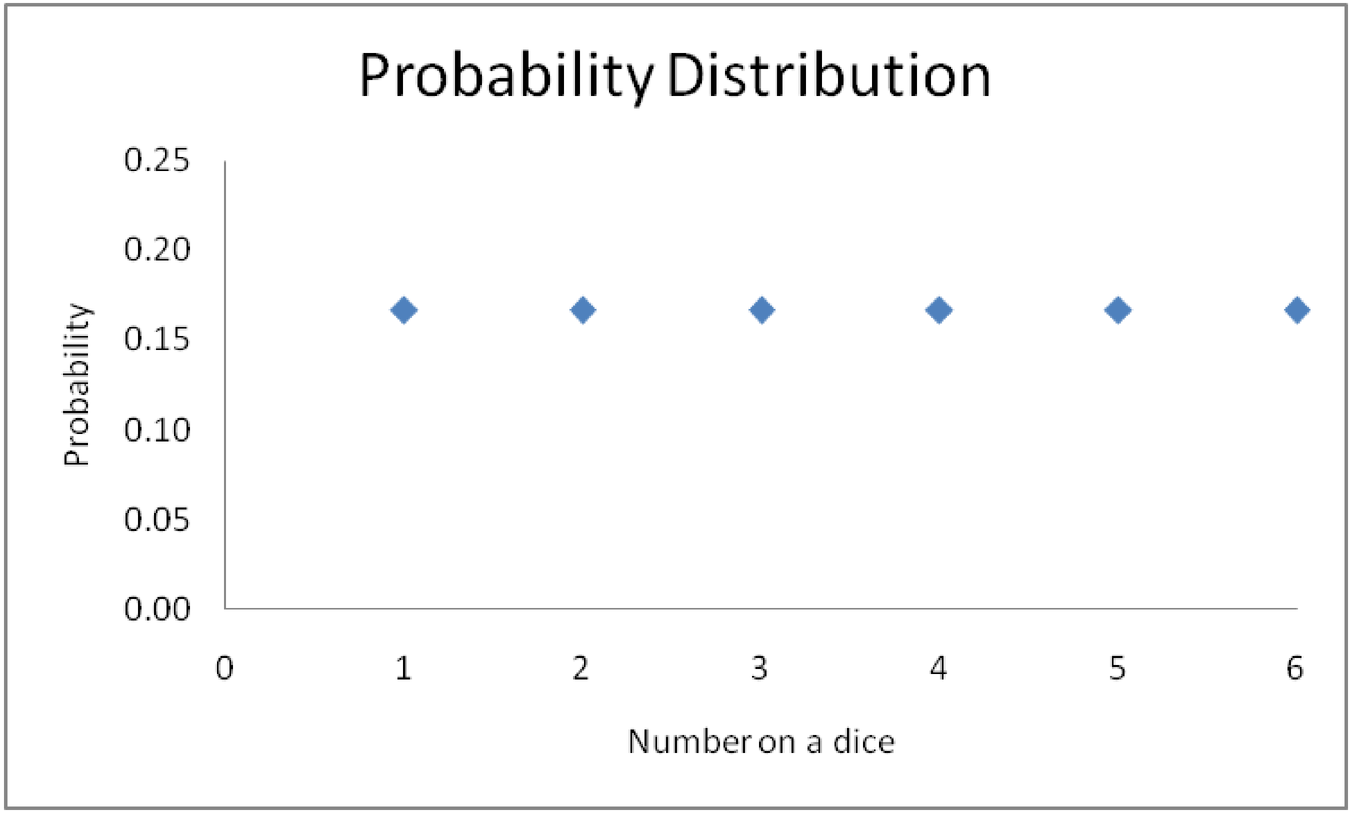
In both the case we must first determine all the possible outcomes for the observations.

Secondly we must determine the probability associated with each of these outcomes or each outcome range.

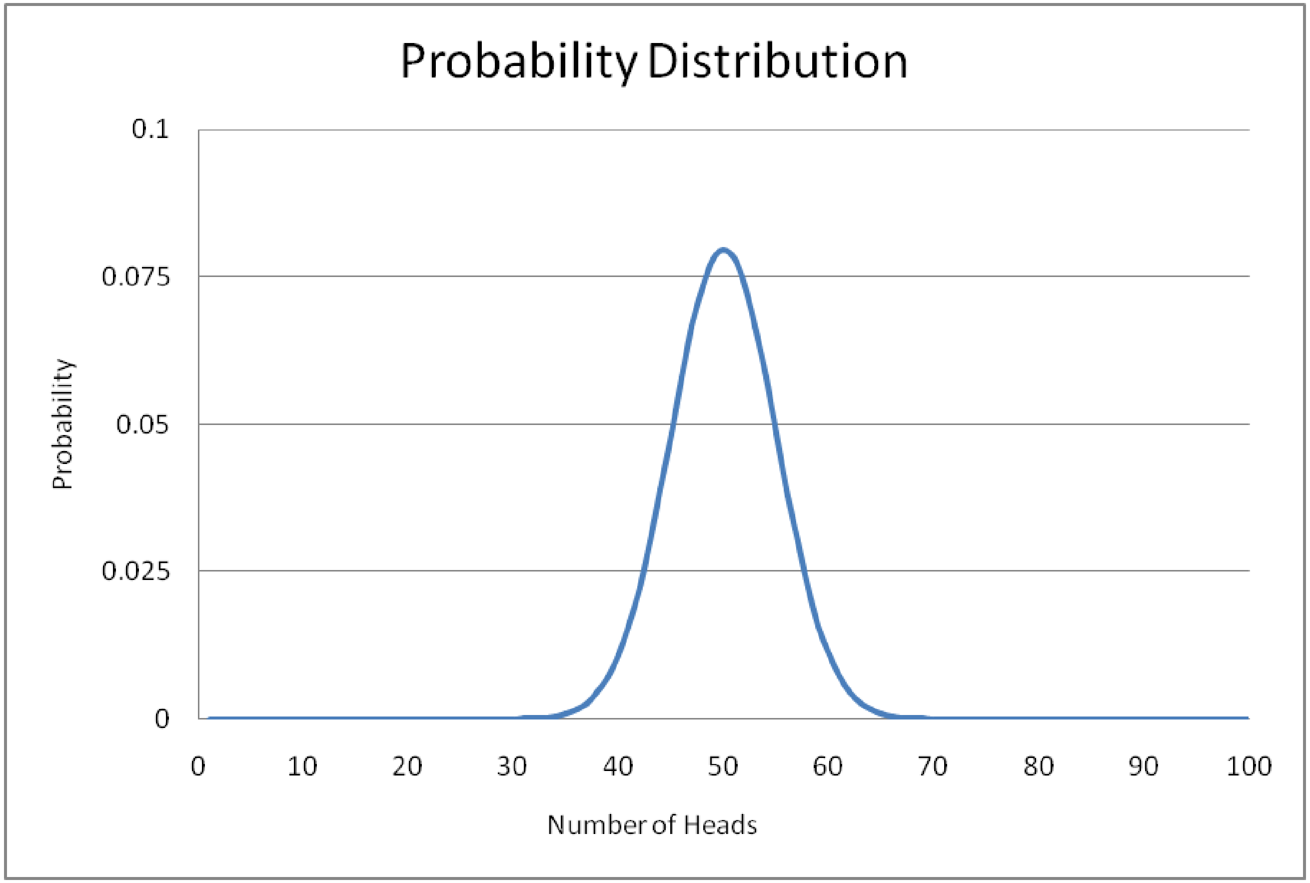
It will follow the following rules

$0 \leq P(x) \leq 1$ (Valid probabilities) and $\sum_{x=\text{all possible outcomes}} P(x) = 1$

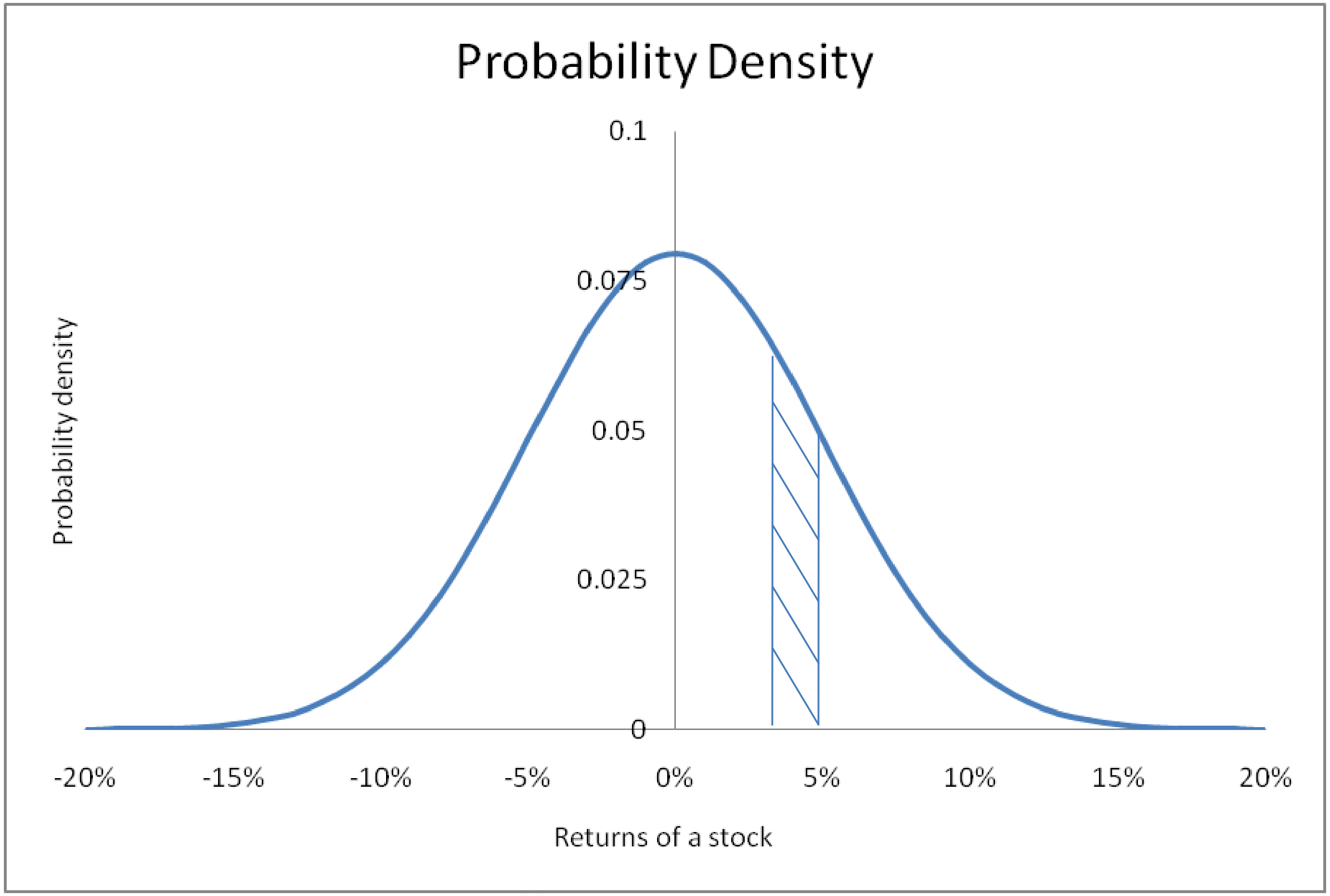
Probability Distribution – Roll a dice



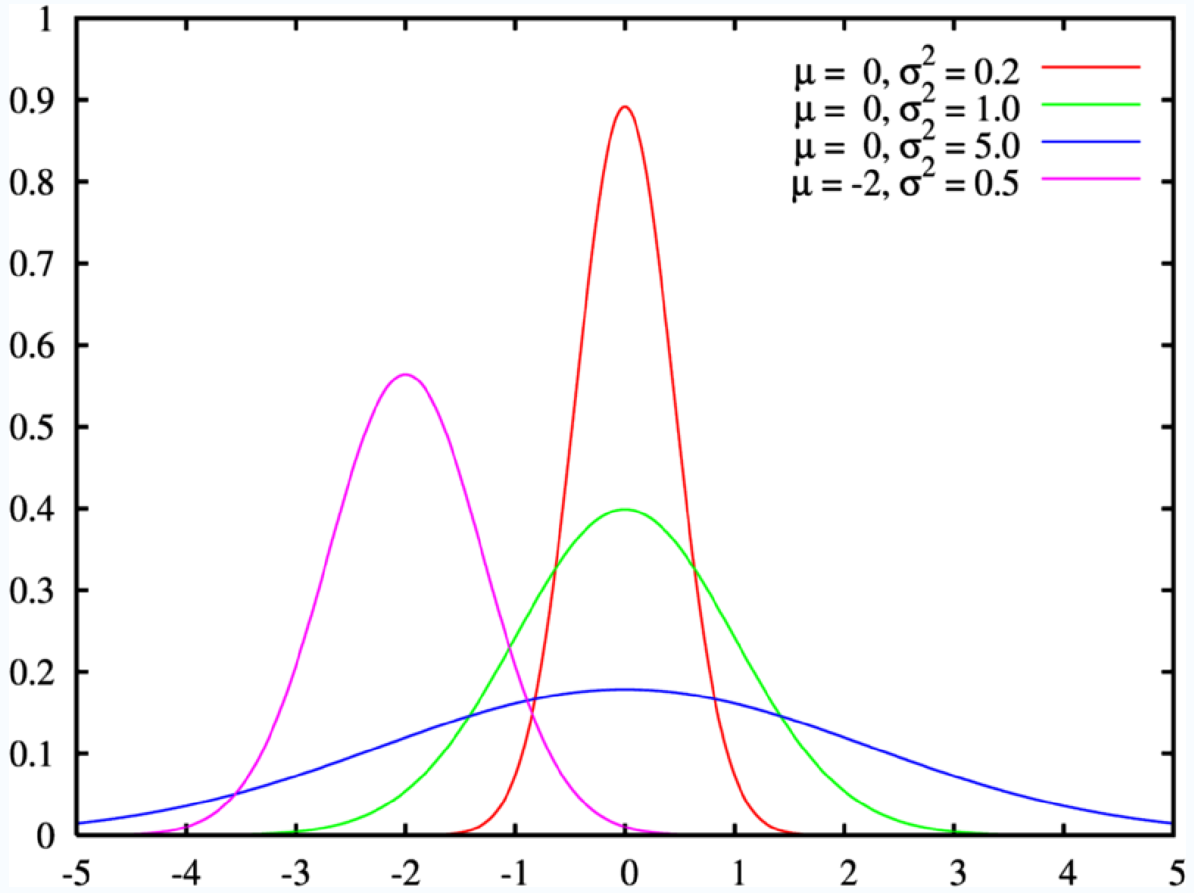
Probability Distribution – Toss a coin



Probability Density – Return of a stock

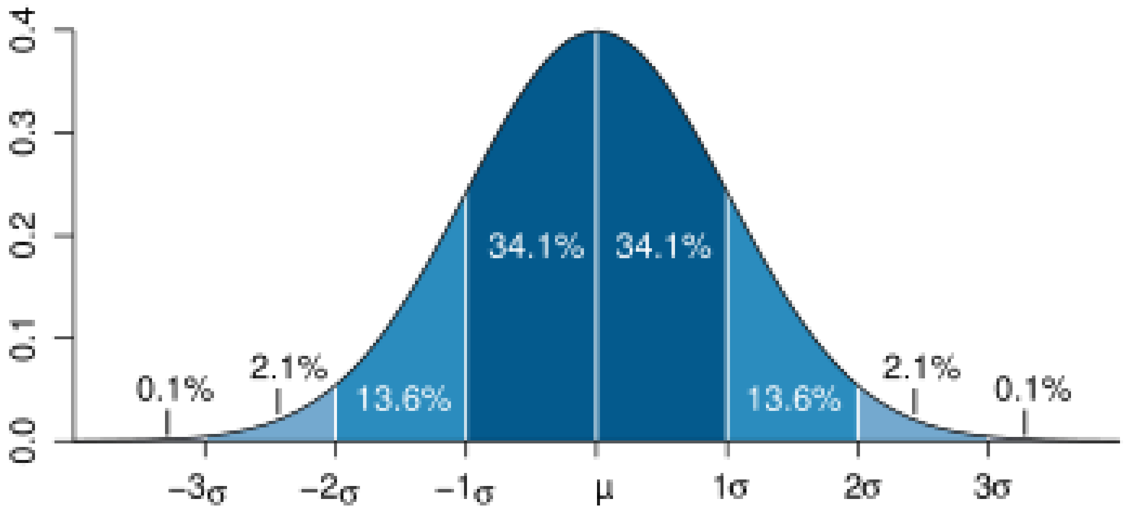


Probability Density Function - Normal Densities



The scale and location (on the horizontal axis) depend on μ and σ. The shape of the distribution is always the same. (Bell curve)

The Empirical Rule and the Normal Distribution



Dark blue is less than one standard deviation from the mean. For the normal distribution, this accounts for about 68% of the set (dark blue) while two standard deviations from the mean (medium and dark blue) account for about 95% and three standard deviations (light, medium, and dark blue) account for about 99.7%.

Expected Values with probability density functions

$$E[X] = \mu = \sum_{i=\text{all outcomes}} P(X_i)\, X_i$$

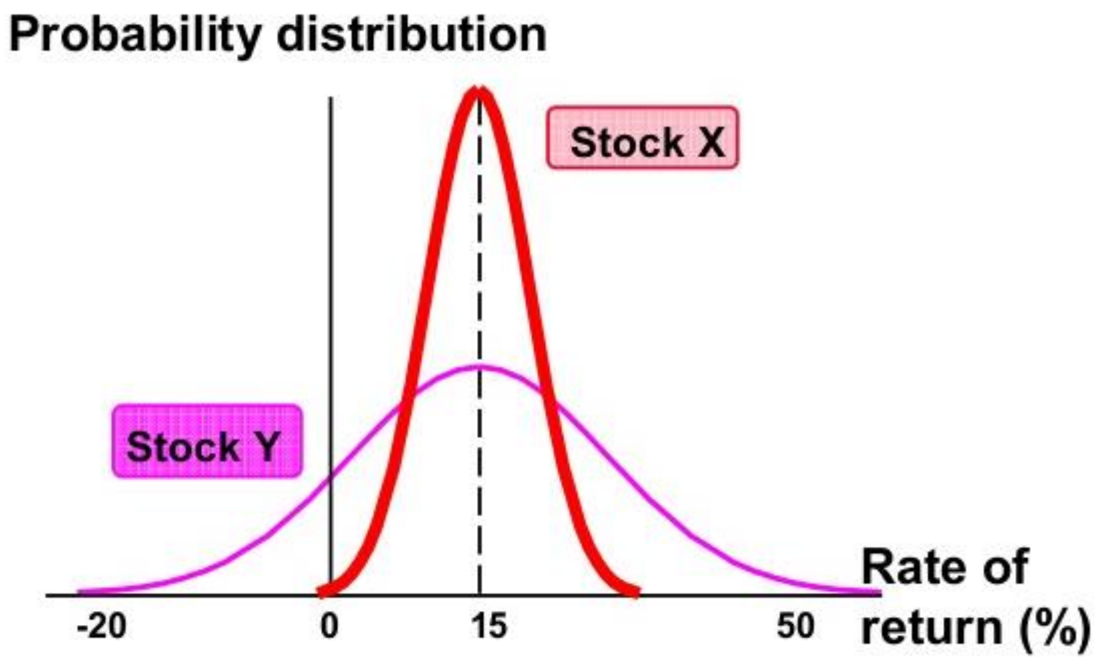
$$\text{Variance} = E[X - \mu]^2 = \sigma^2$$

$$\sigma^2 = \sum_{i=\text{all outcomes}} P(X_i)\,(X_i - \mu)^2$$

where

1. P(x) is the probability density function of variable x
2. $0 \leq P(x) \leq 1$ (Valid probabilities)
3. $\sum_{x=\text{all possible outcomes}} P(x) = 1$

Variance as a measure or risk



- **Which stock is riskier? Why?**

Q & A

Basic Statistics

Relationship between 2 Variables

Linear Regression Models

Regression Modeling

- Regression relationship
  $Y_i = \alpha + \beta x_i + \varepsilon_i$
- Random ε_i implies random Y_i
- Observed random Y_i has two unobserved components:
- Explained: $\alpha + \beta x_i$
- Unexplained: ε_i
- Random component ε_i zero mean, standard deviation σ, normal distribution.

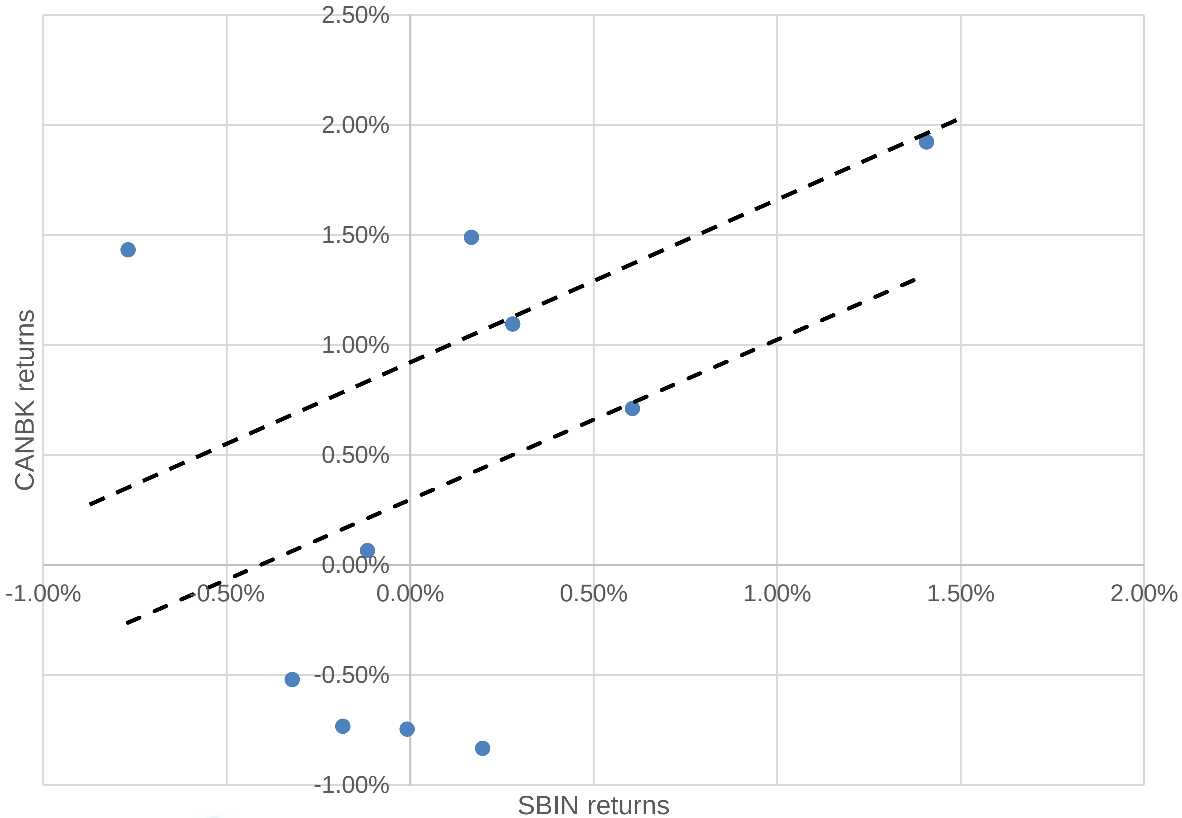
Least squares results

- Regression model
- Sample statistics
- Estimates of population parameters
- How good is the model?
- In the abstract
- Statistical measures of model fit
- Assessing the validity of the relationship

Linear Regression

Date	SBIN	CANBK
15-Nov-22	600.85	309.3
16-Nov-22	599.75	307.05
17-Nov-22	599.05	307.25
18-Nov-22	602.7	309.45
21-Nov-22	598.1	313.95
22-Nov-22	599.1	318.7
23-Nov-22	607.65	324.95
24-Nov-22	609.35	328.55
25-Nov-22	607.4	326.85
28-Nov-22	608.6	324.15
29-Nov-22	608.55	321.75

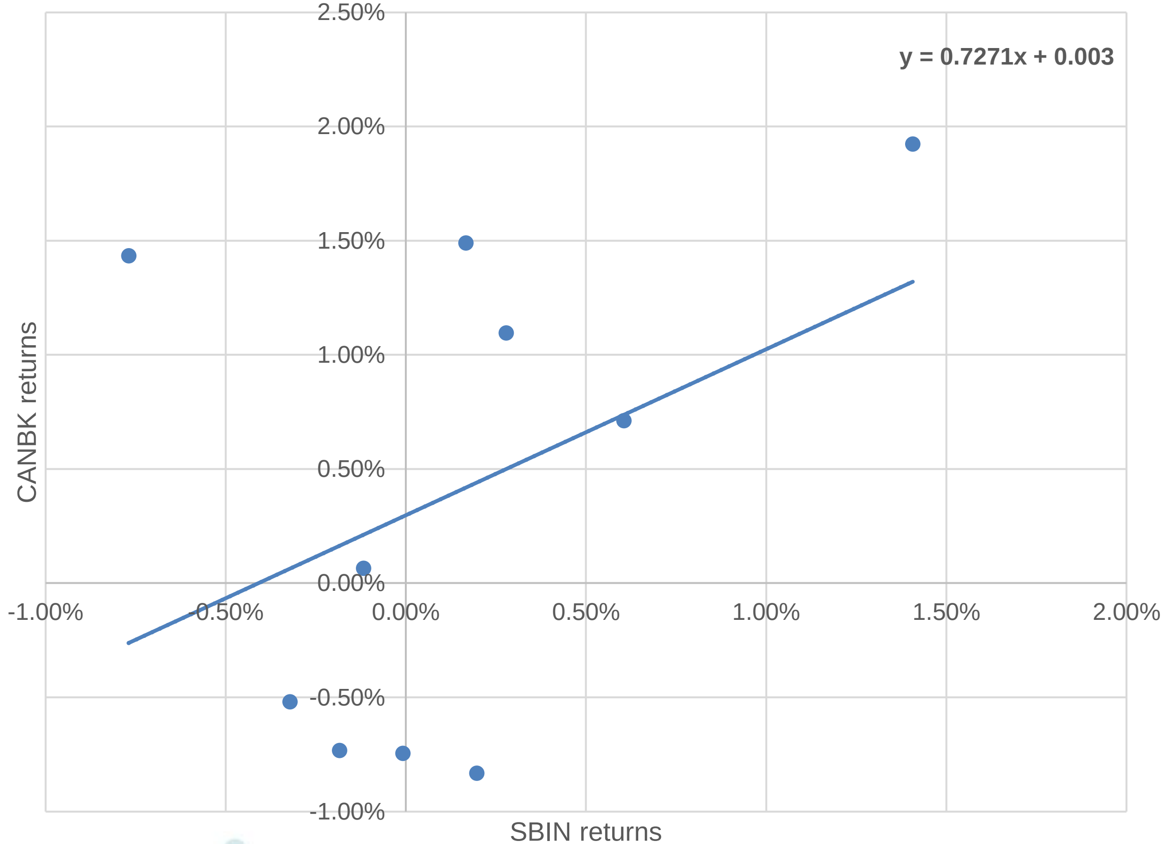
Date	SBIN	CANBK
15-Nov-22	-0.18%	-0.73%
16-Nov-22	-0.12%	0.07%
17-Nov-22	0.61%	0.71%
18-Nov-22	-0.77%	1.43%
21-Nov-22	0.17%	1.49%
22-Nov-22	1.41%	1.92%
23-Nov-22	0.28%	1.10%
24-Nov-22	-0.32%	-0.52%
25-Nov-22	0.20%	-0.83%
28-Nov-22	-0.01%	-0.75%



Linear Regression – Best Fit Line

Date	SBIN	CANBK
15-Nov-22	600.85	309.3
16-Nov-22	599.75	307.05
17-Nov-22	599.05	307.25
18-Nov-22	602.7	309.45
21-Nov-22	598.1	313.95
22-Nov-22	599.1	318.7
23-Nov-22	607.65	324.95
24-Nov-22	609.35	328.55
25-Nov-22	607.4	326.85
28-Nov-22	608.6	324.15
29-Nov-22	608.55	321.75

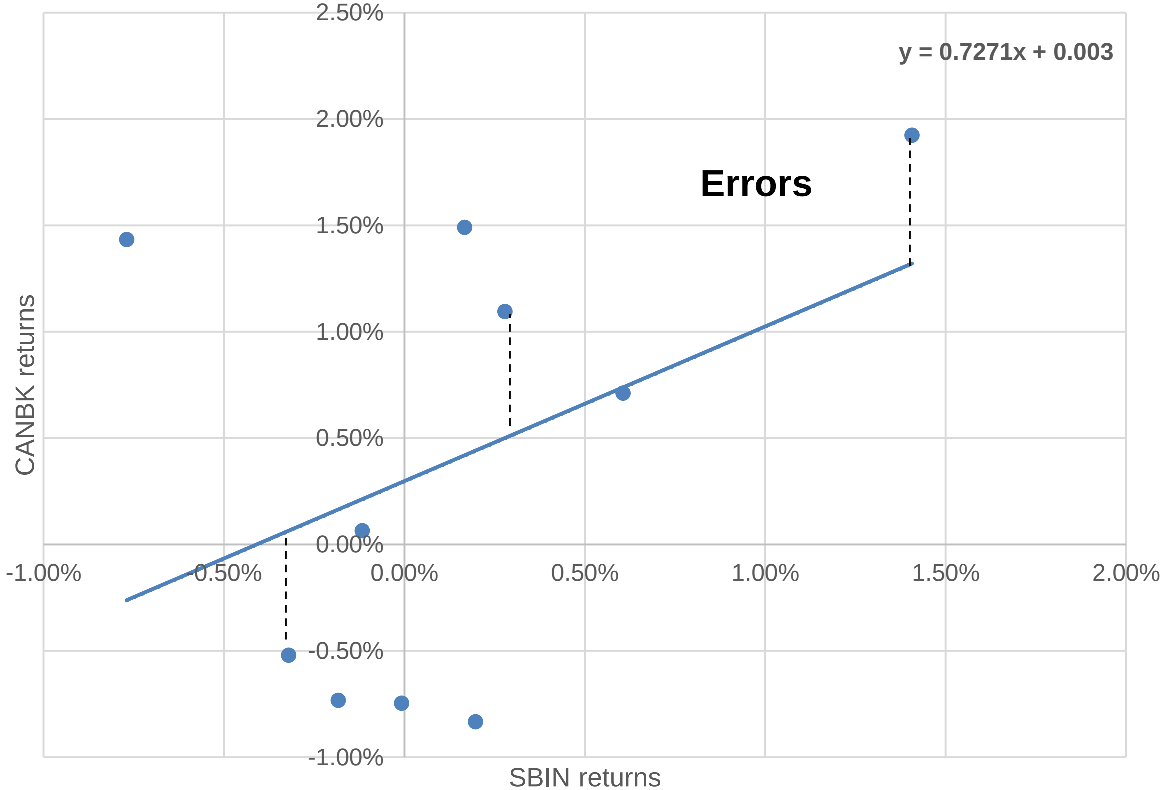
Date	SBIN	CANBK
15-Nov-22	-0.18%	-0.73%
16-Nov-22	-0.12%	0.07%
17-Nov-22	0.61%	0.71%
18-Nov-22	-0.77%	1.43%
21-Nov-22	0.17%	1.49%
22-Nov-22	1.41%	1.92%
23-Nov-22	0.28%	1.10%
24-Nov-22	-0.32%	-0.52%
25-Nov-22	0.20%	-0.83%
28-Nov-22	-0.01%	-0.75%



Linear Regression – Best Fit Line

Date	SBIN	CANBK
15-Nov-22	600.85	309.3
16-Nov-22	599.75	307.05
17-Nov-22	599.05	307.25
18-Nov-22	602.7	309.45
21-Nov-22	598.1	313.95
22-Nov-22	599.1	318.7
23-Nov-22	607.65	324.95
24-Nov-22	609.35	328.55
25-Nov-22	607.4	326.85
28-Nov-22	608.6	324.15
29-Nov-22	608.55	321.75

Date	SBIN	CANBK
15-Nov-22	-0.18%	-0.73%
16-Nov-22	-0.12%	0.07%
17-Nov-22	0.61%	0.71%
18-Nov-22	-0.77%	1.43%
21-Nov-22	0.17%	1.49%
22-Nov-22	1.41%	1.92%
23-Nov-22	0.28%	1.10%
24-Nov-22	-0.32%	-0.52%
25-Nov-22	0.20%	-0.83%
28-Nov-22	-0.01%	-0.75%



Using the Regression Model

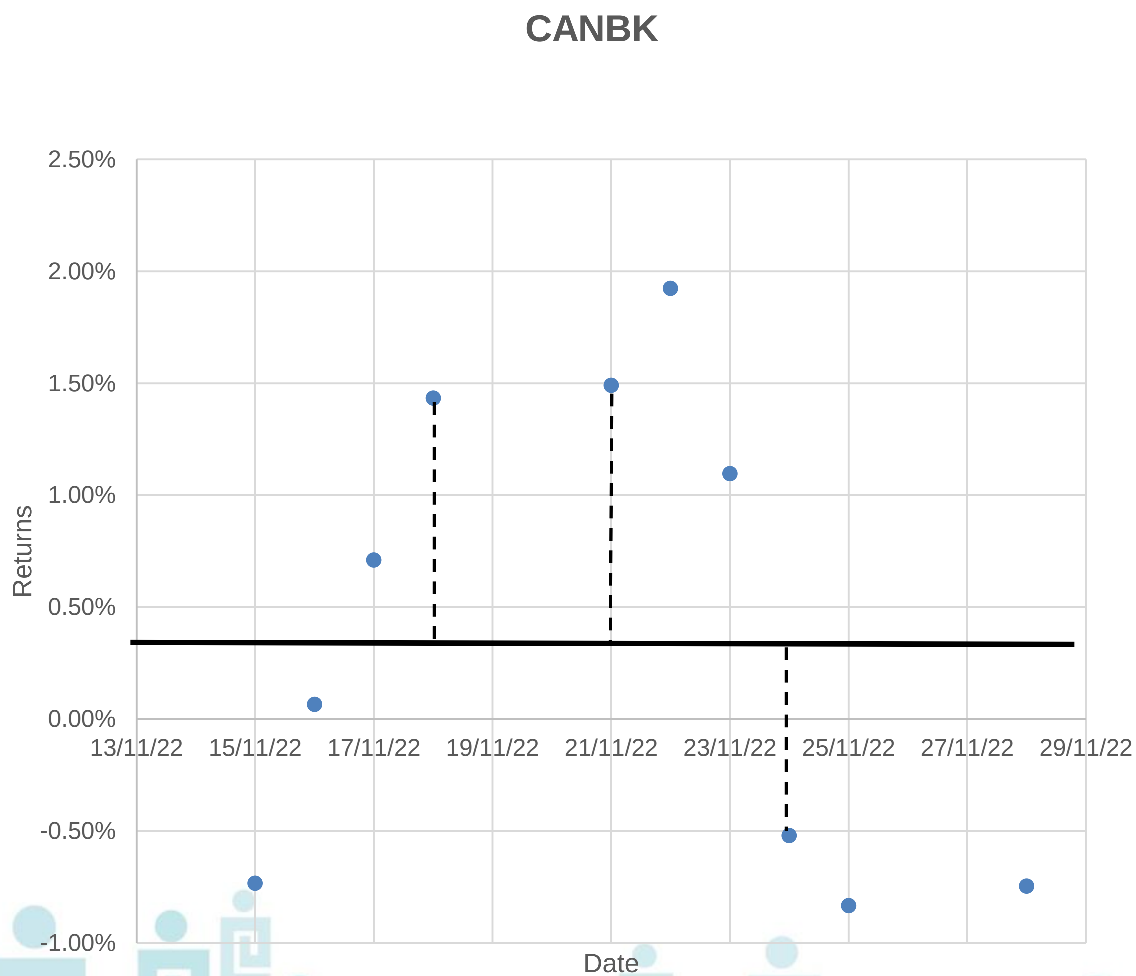
- Prediction: Use x_i as information to predict y_i.
  - The natural predictor is the mean, $\bar{y}$
  - x_i provides more information.

- What we really want to predict is $y_i - \bar{y}$
  - Explain why y_i is different from its mean:
    (1) Because x_i is away from its mean
    (2) Noise
  - The natural predictor is zero
  - Use x_i and the regression to improve.

Linear Regression – Average returns

Date	SBIN	CANBK
15-Nov-22	600.85	309.3
16-Nov-22	599.75	307.05
17-Nov-22	599.05	307.25
18-Nov-22	602.7	309.45
21-Nov-22	598.1	313.95
22-Nov-22	599.1	318.7
23-Nov-22	607.65	324.95
24-Nov-22	609.35	328.55
25-Nov-22	607.4	326.85
28-Nov-22	608.6	324.15
29-Nov-22	608.55	321.75

Date	SBIN	CANBK
15-Nov-22	-0.18%	-0.73%
16-Nov-22	-0.12%	0.07%
17-Nov-22	0.61%	0.71%
18-Nov-22	-0.77%	1.43%
21-Nov-22	0.17%	1.49%
22-Nov-22	1.41%	1.92%
23-Nov-22	0.28%	1.10%
24-Nov-22	-0.32%	-0.52%
25-Nov-22	0.20%	-0.83%
28-Nov-22	-0.01%	-0.75%



R Squared

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

Source	Sum of Squares
Regression	$\Sigma_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
Residual	$\Sigma_{i=1}^{n}e_i^{\,2}$
Total	$\Sigma_{i=1}^{n}(y_i - \bar{y})^2$

Basic Statistics

Application3- The slope or the Beta of a Stock

Beta of the Stock

- Beta is a measure of a stock's volatility in relation to the market.
- By definition, the market has a beta of 1.0, and individual stocks are ranked according to how much they deviate from the market.
- A statistical estimate of beta is calculated by a regression method. For a given asset and a benchmark, the goal is to find an approximate formula:

$$r_a \approx \alpha + \beta r_b$$

where ra is the return of the asset, alpha is the active return and rb is the return of the benchmark.

- Another common expression for Beta is:

$$\beta = \frac{\mathrm{Cov}(r_a, r_b)}{\mathrm{Var}(r_b)}$$

Our example – NIFTY vs ACC

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.650478
R Square	0.423122
Adjusted R Square	0.403892
Standard Error	0.004779
Observations	32

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000503	0.000503	22.00403	5.57E-05
Residual	30	0.000685	2.28E-05		
Total	31	0.001188			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.00073	0.000845	-0.86275	0.395119	-0.00245	0.000997
X Variable 1	0.95963	0.204575	4.690845	5.57E-05	0.541832	1.377429

The Beta of the stock

Q & A

Basic Statistics

Understanding p-value(and the coefficients)

Hypothesis and p-Value

HYPOTHESIS TESTING

Hypothesis testing is the formal use of statistics to determine the probability that a given hypothesis is true and can be accepted or not.

The first and foremost thing to do is to have a hypothesis. This is followed by testing or supporting your hypothesis through data. For this, start with data gathering, followed by the rejection of the hypothesis if the data are inconsistent with it. If the data is consistent with the hypothesis, retain the hypothesis and follow it up with further investigation.

An important point to note is - **Failure to reject is not equivalent to accepting the hypothesis** .

The typical process of Hypothesis Testing consists of the following steps -

1. Formulate the null hypothesis
2. Identify a test statistic that can be used to assess the truth of the null hypothesis.
3. Compute the P-value
4. Compare the -value to an acceptable significance value (α)
5. Based on whether p-value is greater or less than α , accept or reject the null hypothesis

Null Hypothesis

NULL HYPOTHESIS

We start the hypothesis testing by taking a default position. This is the Null Hypothesis.

Commonly, the null hypothesis reflects the fact that the sample observations result purely from chance. In other words - there is no relationship between two measured phenomena, or no difference among groups.

Null Hypothesis is considered **TRUE** until proven otherwise.

The hypothesis contrary to the null hypothesis, usually that the observations are the result of a real effect, is known as the **alternative hypothesis.**

TEST STATISTICS

Test Statistic is a single measure of some attribute of a sample (i.e. a statistic) used in statistical hypothesis testing.

It is defined in such a way as to quantify, within observed data, behaviours that would distinguish the null from the alternative hypothesis.

The test statistic compares your data with what is expected under the null hypothesis. The test statistic is used to calculate the p-value.

The test statistics are changes randomly from one random sample to another.

For example, the test statistic for a Z-test is the Z-statistic, which has the standard normal distribution under the null hypothesis.

Once you have decided on the null hypothesis and have the test statistics, you need a reference to base your decision on. The strength of evidence in support of a null hypothesis is measured by the **P-value**.

P-value is the probability of observing a more extreme value of test statistics than the one calculated, provided the null hypothesis is true. In other words, P-value is the probability of obtaining a result equal to or more than the observed value.

To accept or reject a null hypothesis, we compare the P-value with a value on the test distribution, called the critical value.

Remember, p-value is the probability that we calculate from the observed data.

SIGNIFICANCE LEVEL

Significance level is used to calculate the critical value used to compare the P-value against in order to accept or reject a null hypothesis.

Significance level, denoted by "α" , is a probability threshold below which the null hypothesis is rejected.

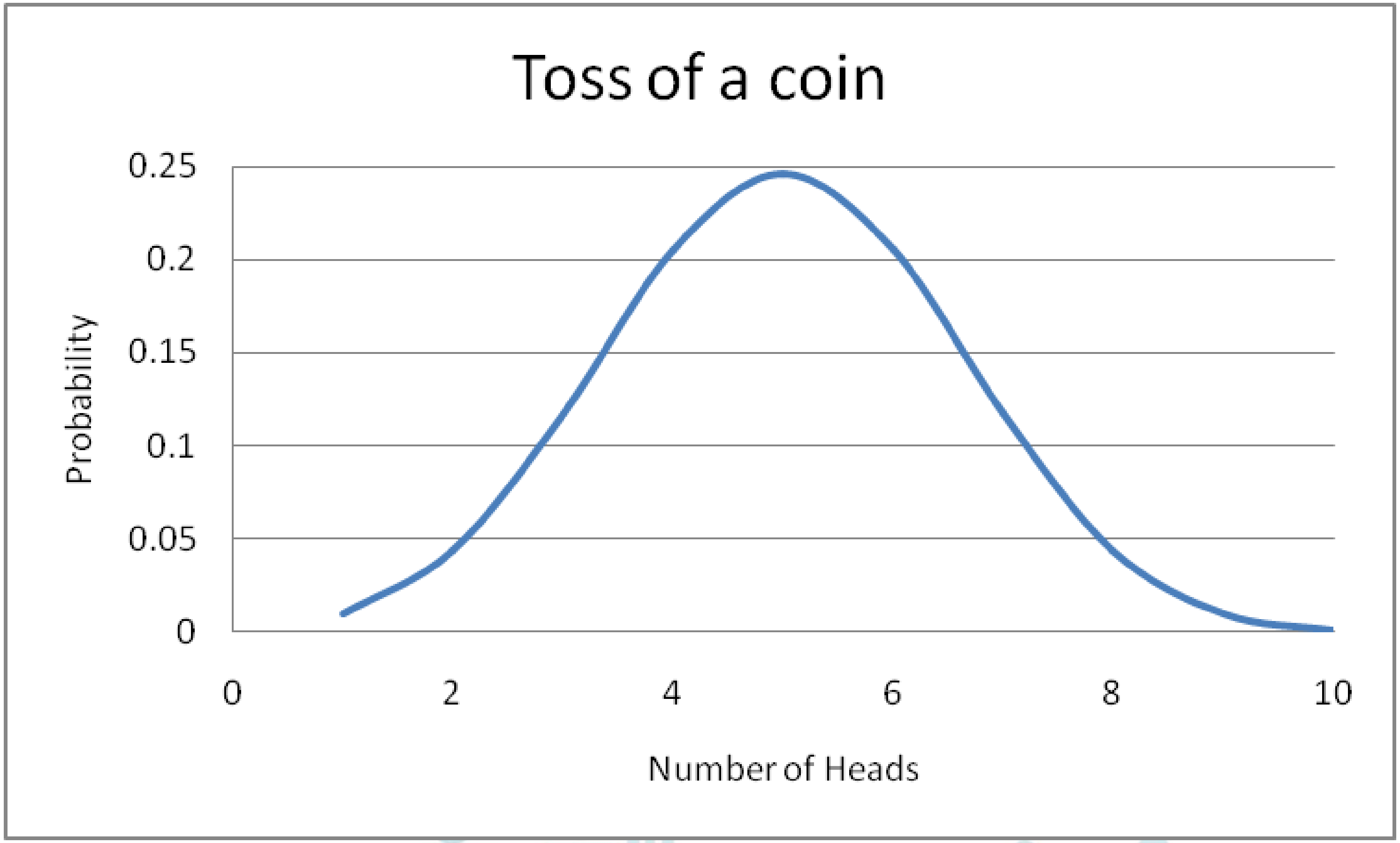
Significance level also denotes the probability of making a TYPE I error. **Type I error** is the false rejection of the null hypothesis.

Remember, that unlike p-value, significance level is not a calculated quantity. Instead, it is arrived at by consensus among the researchers. It is set in advance before the hypothesis testing begins. Conventionally, it is taken to be 1%, 5% or 10% i.e. 0.01, 0.05 or 0.10 respectively.

The final step in hypothesis testing is to compare the calculated p-value with the significance level i.e. - Compare the *P*-value to α.

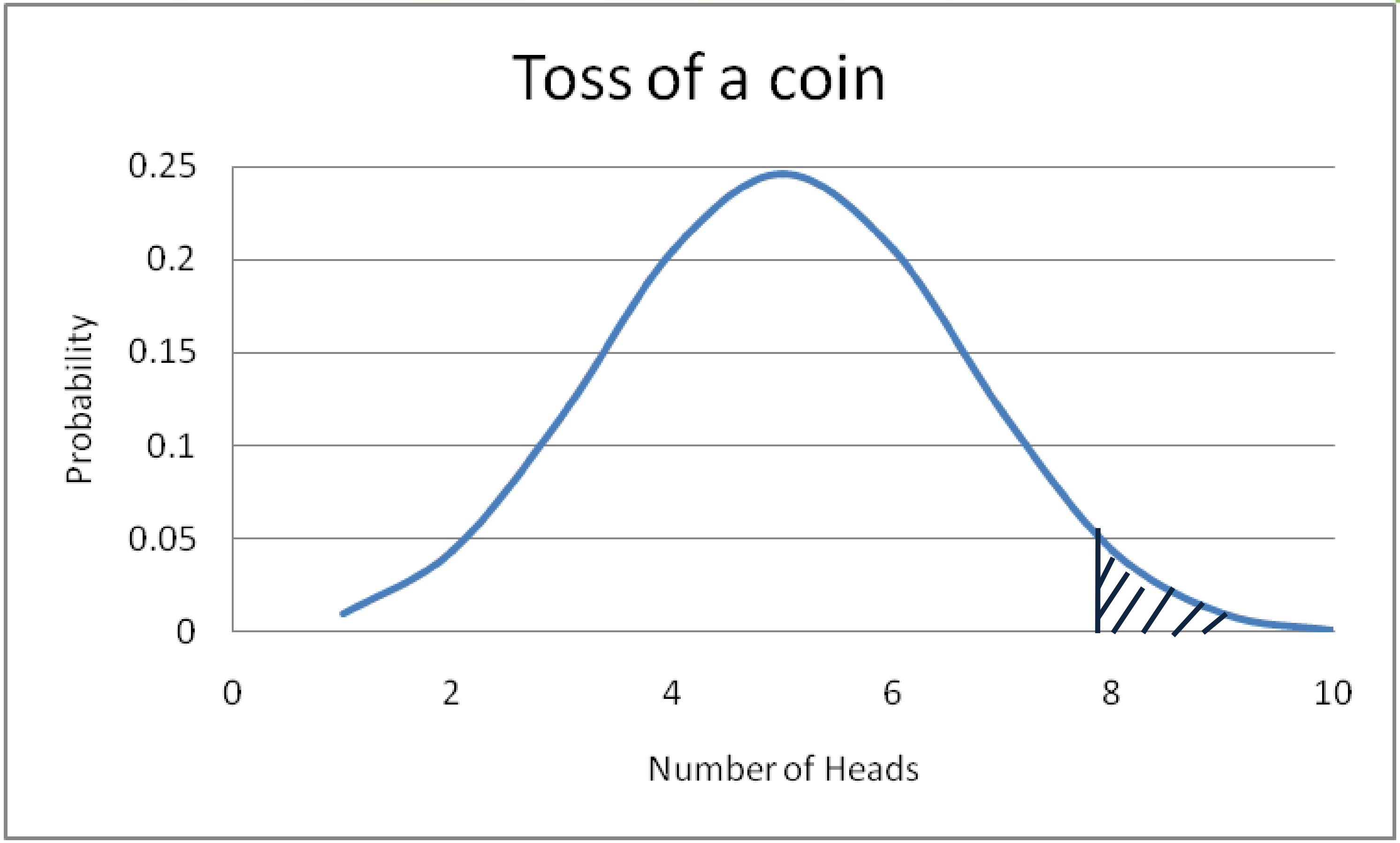
If the *P*-value is less than (or equal to) α, reject the null hypothesis in favor of the alternative hypothesis. If the *P*-value is greater than α, do not reject the null hypothesis.

p-Value



Toss of a coin
0.25
0.2
0.15
0.1
0.05
0
Probability
0
2
4
6
8
10
Number of Heads

p-Value


Toss of a coin
Probability
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10
Number of Heads

What does the p value tell us?

Question: Which of the following statements about *p*-values is correct?

1. A *p*-value of 5% implies that the probability of the null hypothesis being true is (no more than) 5%.

2. A *p*-value of 0.005 implies much more "statistically significant" results than does a *p*-value of 0.05.

3. The *p*-value is the likelihood that the findings are due to chance.

4. A *p*-value of 1% means that there is a 99% chance that the data were sampled from a population that's consistent with the null hypothesis.

5. None of the above.

What does the p value tell us?

<u>Answer</u>: 5.

For the record, let's have a clear statement of what a *p*-value is, and then we can move on.

- *The p-value is the probability of observing a value for our test statistic that is as extreme (or more extreme) than the value we have calculated from our sample data, given that the null hypothesis is true.*

Source: http://davegiles.blogspot.in/2011/04/may-i-show-you-my-collection-of-p.html

Our example – estimate of the slope and the confidence level.

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.00073	0.000845	-0.86275	0.395119	-0.00245	0.000997
X Variable 1	0.95963	0.204575	4.690845	5.57E-05	0.541832	1.377429

The value of coefficient(s)
The value of the Standard Error(s)
The Range(s) for 95% Confidence Level
The p-values for coefficient(s).

Source: http://davegiles.blogspot.in/2011/04/may-i-show-you-my-collection-of-p.html

Basic Statistics

Suggested Exercises for the Participants

Problems to be attempted for Self-study

- Minimum Volatility Portfolio
  - Calculate rolling n[=50/100/250] day variance and return for a universe of 50 stocks.
  - Choose 10 lowest volatility stocks. Allocate equal capital to each of 10 stock. Calculate returns for this portfolio.
  - Compare returns and volatility of the portfolio to index.
- Beta = 1 Portfolio
  - Calculate rolling n[=50/100/250] beta from a universe of 50 stocks
  - Choose 10 lowest/highest beta stocks and assign weights such that the net beta is 1.
  - Compare returns and volatility of the portfolio to index.

Extra – Compute the Drawdown in each case and compare against index buy and hold strategy.

Q & A