Econ 106VL

Statistical and Probability Review

Professor Pierre-Olivier Weill

Outline

- Practical information about the class
- Random variables, probability distributions
- Introduce today's lab assignment

Class times and venues (Lec 01)

- Main lecture
  - F 9:00am-9:50am, Dodd Hall 161

- Computer labs
  - Lab 1A: F 10:00pm-10:50pm, Bunche 3178
  - Lab 1B: F 11:00am-11:50am, Bunche 3178
  - Lab 1C: F 3:00pm-3:50pm, Perloff 1102

Class times and venues (Lec 02)

- Main lecture
  - F 10:00am-10:50am, Dodd 161
- Computer Labs
  - Lab 2A: F 1:00pm-1:50pm, Haines A18
  - Lab 2B: F 2:00pm-2:50pm, Pub Aff 1278
  - Lab 2C: F 4:00pm-4:50pm, Perloff 1102

TA's

- TAs are in charge of the labs, 106VL
- The TAs will hold office hours every week
  - Starting next week

TAs

- Give the lab lecture
- Provide Excel help during the lab section

Basic structure of the Friday lab

- **A lab lecture**
  - Review concepts and tools used in Excel project
  - Provide feedback on previous projects
- **A lab section**
  - Get started with the Excel project
  - It **will** take longer than the 50 minutes lab section
  - **Work in groups of 3**
- **During the subsequent week**
  - Continue working on the project
  - Create a **4-5 slides** Powerpoint presentation
  - Grading: credit/no credit

Guest lectures in 106VL

- We may have a couple of guest speakers
- Be sure to be present!
  - we will take attendance
- I will send you announcements to save dates

The Excel project presentation

- Summarize the findings:
  - key numbers, graphs
  - main take away of the project
- Acceptable format: **pdf or powerpoint**
  - Turning in the spreadsheet gives no credit
- List the group members on the first page
  - Keep the same group starting with the 2nd project
- Turn it in on the website by ***Thursday midnight***

Lab Account

- Students should have a lab account:
- Instructions:

https://computing.sscnet.ucla.edu/labs/lab-accounts/

- Lab accounts can be created by presenting your Student ID or a picture ID to a lab consultant at the Front Desk.
- Accounts are created automatically for students who are enrolled in a class that is using the labs for the quarter
- Login directions are provided by the FrontDesk Consultants and are available on each computer screen
- Your login ID is the first letter of your first and last name combined with the last sixnumbers of your UCLA University ID number. For example, Joe Bruin with StudentID 123456789 will become JB456789. Your temporary password is joebruin. Youwill be prompted to change your password upon logging in the first time.

Statistic and Probability Review

Random Variables

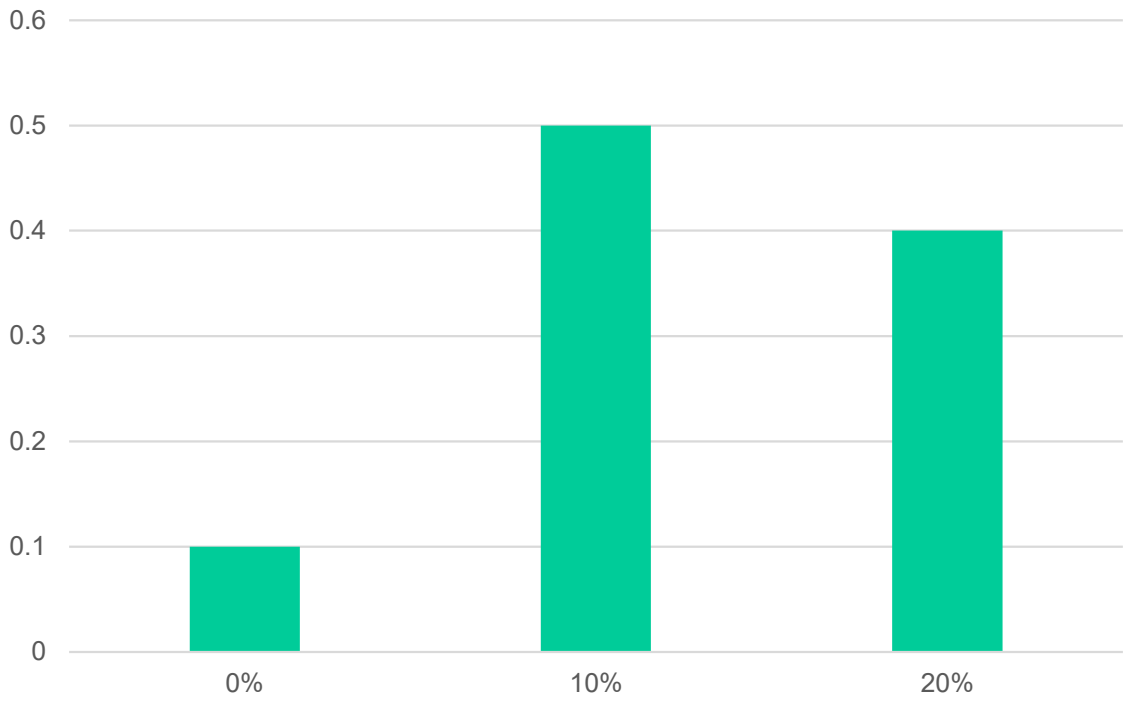
- Definition: a value representing an outcome of an uncertain experiment
- The outcome may be
  - Discrete:
    - "scenario analysis"
      - Value depend on a discrete number of "scenarios"
      - Example: outcome of a soccer game
      - Example : 3 possible business cycle conditions
      - ✓ {recession, normal, boom}
      - Example : 2 possible flips of a coin
      - ✓ {heads, tails}
  - Continuous:
    - A continuum of scenarios
      - Example: the weight of a baby
      - Example: the return of a stock

Probability Distribution

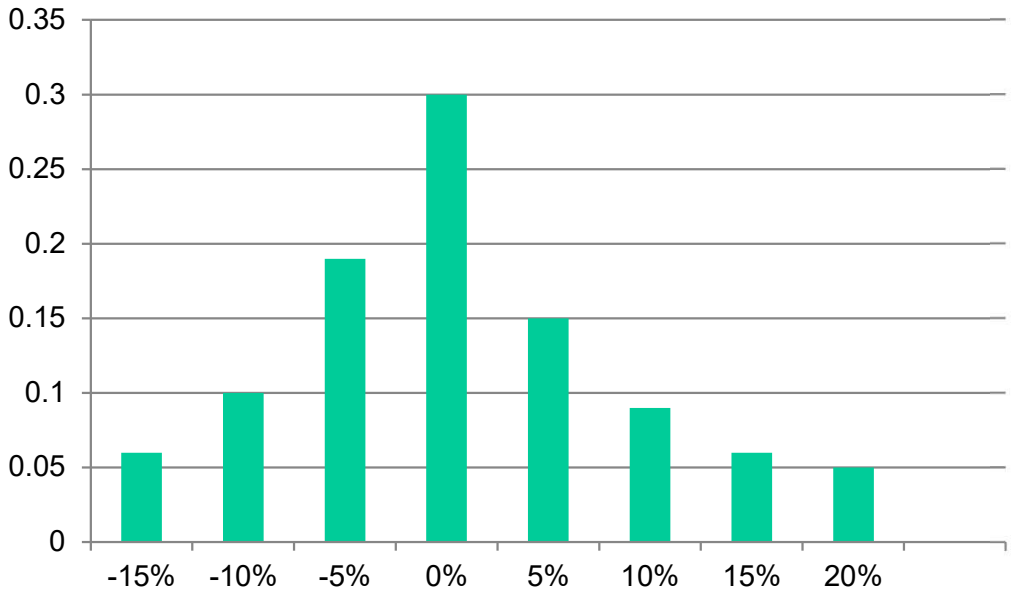
- The likelihood of each possible event(a set of outcomes)
- Example of a discrete outcome
  - Fair coin: 50% head, 50% tail
  - Rigged coin: 60% head, 40% tail
  - For continuous outcome
    - Normal distribution

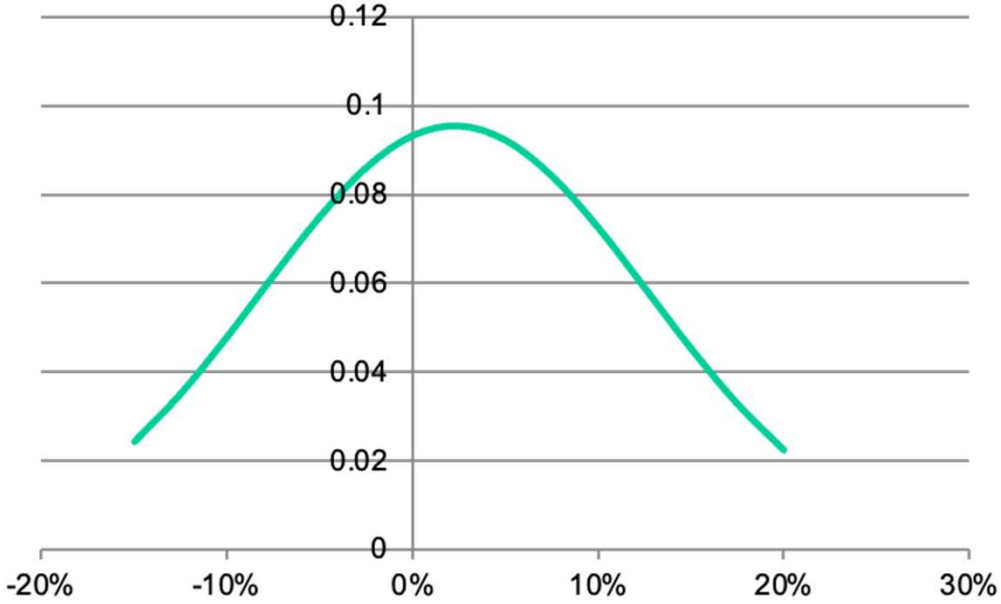
Example: Discrete Probability Distributions

Scenario	Probability	Return
Market Crash	.1	0%
Stable Markets	.5	10%
Market Rally	.4	20%



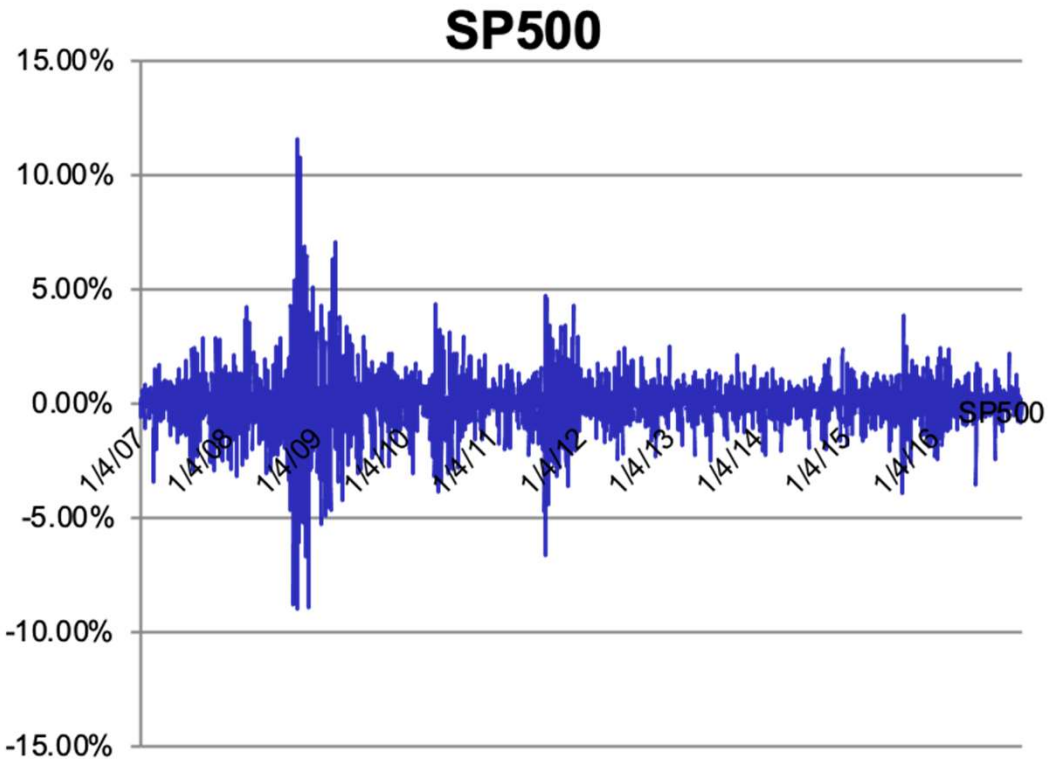
Example:
Continuous Probability Distribution





Example

- Daily S&P500 daily returns, 2007-2017



A snapshot of the data

	A	B	C
1	Date	Adj Close	Return SP 500
2	9/20/12	1460.26	-0.05%
3	9/19/12	1461.05	0.12%
4	9/18/12	1459.32	-0.13%
5	9/17/12	1461.19	-0.31%
6	9/14/12	1465.77	0.40%
7	9/13/12	1459.99	1.63%
8	9/12/12	1436.56	0.21%
9	9/11/12	1433.56	0.31%
10	9/10/12	1429.08	-0.61%
11	9/7/12	1437.92	0.40%
12	9/6/12	1432.12	2.04%
13	9/5/12	1403.44	-0.11%
14	9/4/12	1404.94	-0.12%
15	8/31/12	1406.58	0.51%
16	8/30/12	1399.48	-0.78%
17	8/29/12	1410.49	0.08%
18	8/28/12	1409.3	-0.08%
19	8/27/12	1410.44	-0.05%
20	8/24/12	1411.13	0.65%
21	8/23/12	1402.08	-0.81%
22	8/22/12	1413.49	0.02%
23	8/21/12	1413.17	-0.35%
24	8/20/12	1418.13	0.00%
25	8/17/12	1418.16	0.19%

Count observation **below** each threshold

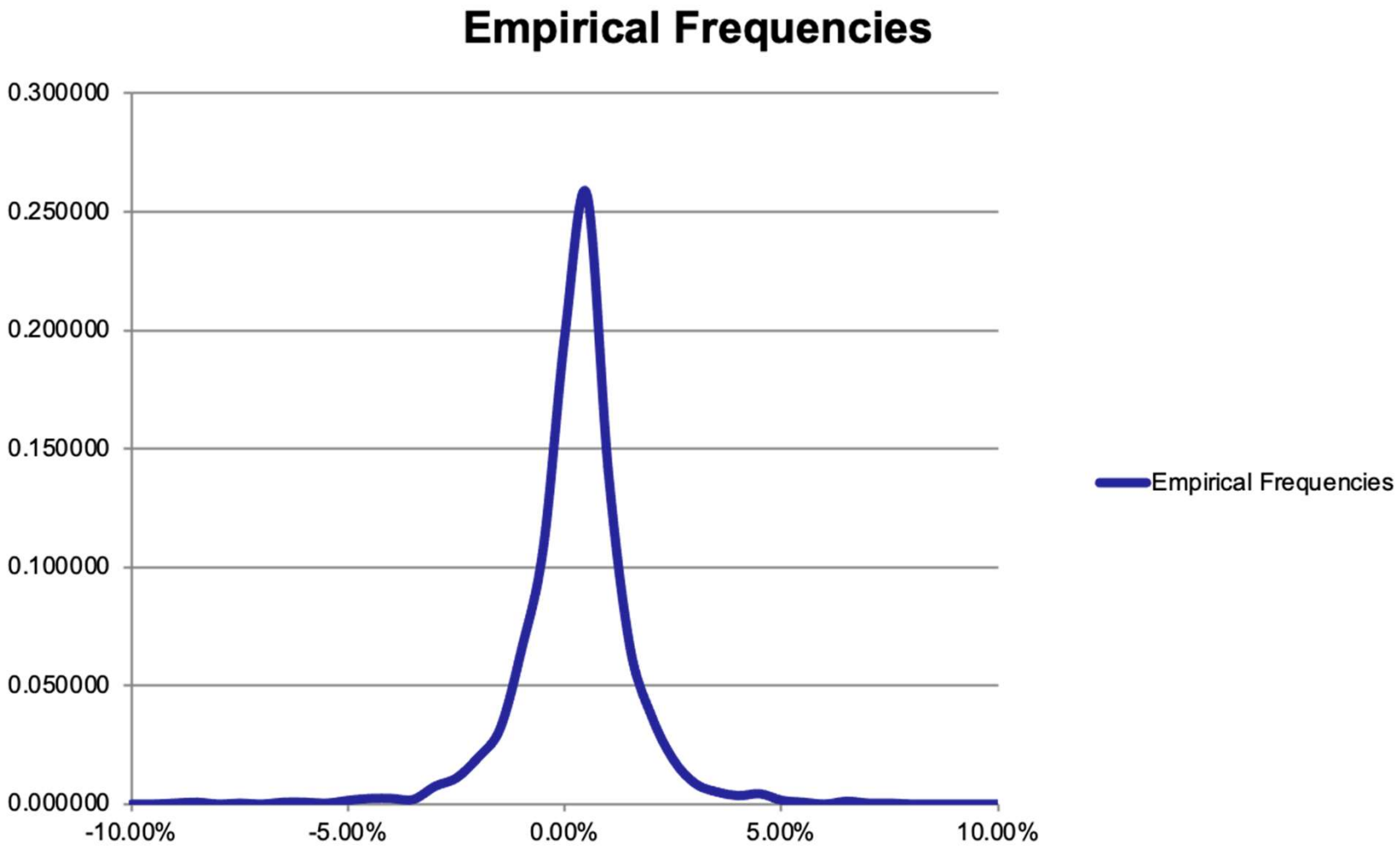
=FREQUENCY($C$2:$C$2520,H3)

Calculate frequency of observation inside the bin

(I4-I3)/2517

Create bins

	H	I	J
		Data	
	Bins	Count below	Frequency
	-10.00%	0.000000	0.000000
	-9.50%	0.000000	0.000000
	-9.00%	1.000000	0.000397
	-8.50%	3.000000	0.000795
	-8.00%	3.000000	0.000000
	-7.50%	4.000000	0.000397
	-7.00%	4.000000	0.000000
	-6.50%	6.000000	0.000795
	-6.00%	8.000000	0.000795
	-5.50%	9.000000	0.000397
	-5.00%	13.000000	0.001589
	-4.50%	19.000000	0.002384
	-4.00%	25.000000	0.002384
	-3.50%	30.000000	0.001986
	-3.00%	49.000000	0.007549
	-2.50%	77.000000	0.011124


Empirical Frequencies
0.300000
0.250000
0.200000
0.150000
0.100000
0.050000
0.000000
-10.00%
-5.00%
0.00%
5.00%
10.00%
Empirical Frequencies

Expected Value

the probability-weighted average of the possible outcomes

- Suppose the return R_i on an asset *i* is equal to R_i *(s)* with probability *p(s)* for *s=1,…,S.*
- Then the expected return is:

$$\mathbb{E}[R_i] = Proba(1) \times R_i(1) + Proba(2) \times R_i(2) + \ldots + Proba(S) \times R_i(S)$$

$$= \sum_{s=1}^{S} Proba(s) R_i(s)$$

- The expected value is the average outcome if the event was repeated infinitely often.

Example: Expected Value

Scenario	Probability	Return
Market crash	.1	0%
stable markets	.5	10%
market boom	.4	20%

$$\begin{aligned}\mathbb{E}(R) = & R_{\text{market crash}} \times Proba_{\text{market crash}} \\ & + R_{\text{stable market}} \times Proba_{\text{stable market}} \\ & + R_{\text{market boom}} \times Proba_{\text{market boom}}\end{aligned}$$

- E(R) = 0%*0.1+10%*0.5+20%*0.4=13%

Variance

the average squared deviation from the expected value

$$\sigma_i^2 = \mathbb{E}\left[(R_i(s) - \mathbb{E}[R_i])^2\right]$$

$$= \sum_{s=1}^{S} Proba(s) \times (R_i(s) - \mathbb{E}[R_i])^2$$

- The *standard deviation* (SD) is the square root of the variance:

$$\sigma_i = \sqrt{\sigma_i^2}$$

- ‘volatility’ is another word for ‘standard deviation’.

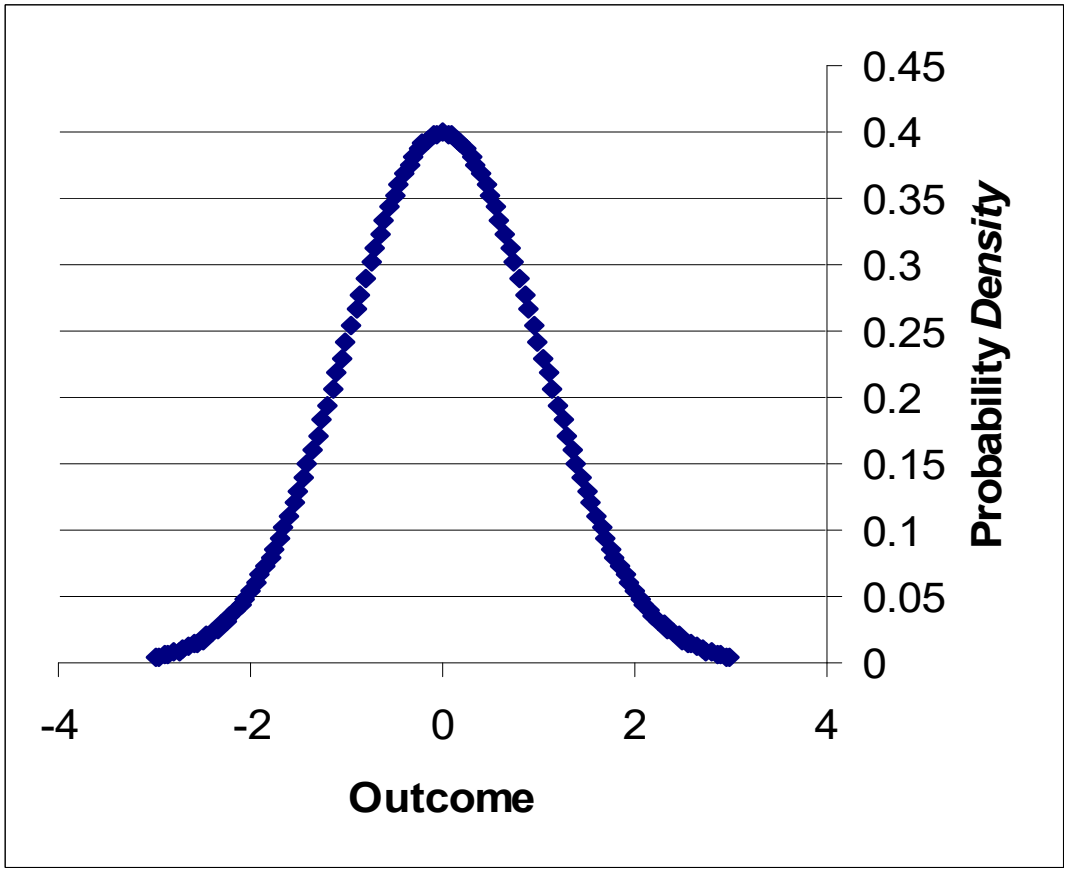
Example Variance

Scenario	Probability	Return
Market Crash	.1	0%
Stable Markets	.5	10%
Market Rally	.4	20%

- $V(R) = 0.1*(0\%-13\%)^2 + 0.5*(10\%-13\%)^2 + 0.4*(20\%-13\%)^2 = .0041$
- Standard deviation

  = square root of .0109 =6.4%

Normal Distribution (Bell Curve)



Summarized by its mean and standard deviation

Covariance

- The *covariance* between two random variables is the average of the products of their deviations from the mean:

$$cov(R_i, R_j) = \mathbb{E}\left[\left(R_i - \mathbb{E}\left[R_i\right]\right)\left(R_j - \mathbb{E}\left[R_j\right]\right)\right]$$

$$= \sum_{s=1}^{S} Proba(s) \times \left(R_i(s) - \mathbb{E}\left[R_i\right]\right)\left(R_j(s) - \mathbb{E}\left[R_j\right]\right)$$

- The covariance is
  - **Positive** if the random variables tend to be above (below) their means *at the same time*
  - **Negative** if the one variable tends to be above (below) its mean when the other is below (above) its mean

Covariance

	Proba	HPR of AAPL	HPR of XLU
Market Crash	1/3	0%	0%
Stable Markets	1/3	10%	20%
Market Rally	1/3	20%	40%

Mean of 1 =10%

Variance of 1 =.0067

Mean of 2 =20%

Variance of 2 =.0267

Covariance of 1 and 2 =0.01333

Correlation

- The *correlation* is the covariance between two random variables, divided by their standard deviations:

$$\rho_{ij} = \frac{\mathrm{cov}(R_i, R_j)}{\sigma_i \sigma_j}$$

- The correlation is scaled so that:

$$-1 \leq \rho_{ij} \leq 1$$

Estimating Mean, Variance, and Covariance from Historical Data

- Use the "sample counterpart" of the definition:

$$\hat{E}(R_i) = \frac{1}{T}\sum_{t=1}^{T} R_i(t)$$

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T} [R_i(t) - \hat{E}(R_i)]^2$$

$$\hat{\mathrm{cov}}(R_i, R_j) = \frac{1}{T-1}\sum_{t=1}^{T} [R_i(t) - \hat{E}(R_i)][R_j(t) - \hat{E}(R_j)]$$

Lab Assignment, Week 1

- Exercise 1: Basic Descriptive Statistics
  - Download monthly historical prices of a publically traded firm and the S&P 500 (^GSPC) from Yahoo! Finance.
    - Data contains the Open, Close, High, Low, Volume, and Adjusted Close
  - Calculate Returns
    - R(t) = Adj. Close(t)/Adj. close (t-1) - 1
  - Calculate mean & std. dev. of returns
    - AVERAGE() & STDEVP() functions
  - Calculate Correlation of 2 series & scatter plot
    - CORREL()

Lab Assignment, Week 1

- Exercise 2: Do daily returns follow a normal distribution?
  - Download 10 years of daily price data for the S&P 500
  - Calculate means & std. dev. of returns
  - Create a Frequency table for the data
    - Use Excel frequency function
  - Create a Frequency table for normal dist.
    - More on this in your lab section
  - Plot & compare the two frequency tables