金程教育
GOLDEN FUTURE
可 信 赖 的 财 经 培 训 专 家

CFA一级培训项目

Quantitative Methods

单晨玮
金程教育高级培训师
地点： ■ 上海 □北京 □深圳

Topic Weightings in CFA Level I

Session NO.	Content	Weightings
Study Session 1	Ethics & Professional Standards	15
Study Session 2-3	Quantitative Analysis	12
Study Session 4-6	Economics	10
Study Session 7-10	Financial Reporting and Analysis	20
Study Session 11	Corporate Finance	8
Study Session 12	Portfolio Management and Wealth Planning	5
Study Session 13-14	Equity Investment	10
Study Session 15-16	Fixed Income	12
Study Session 17	Derivatives	5
Study Session 18	Alternative Investments	3

Quantitative Methods

- Time Value Calculation
  - R5 The Time Value of Money
  - R6 Discounted Cash Flow Applications
- Probability & Statistics
  - R7 Statistical Concepts and Market Returns
  - R8 Probability Concepts
  - R9 Common Probability Distributions
- Inferential statistics
  - R10 Sampling and Estimation
  - R11 Hypothesis Testing

R5 Time Value of Money

- Time Value of Money
  1. Required interest rate on a security的组成
  2. EAR
  3. Annuities的计算：FV, PV, required payment

R5 Time Value of Money

- **Required rate of return** is
  - affected by the supply and demand of funds in the market;
  - the return that investors and savers require to get them to willingly lend their funds;
  - usually for particular investment.
- **Discount rate** is
  - the interest rate we use to discount payments to be made in the future.
  - usually used interchangeably with the interest rate.
- **Opportunity cost is**
  - also understood as a form of interest rate. It is the value that investors forgo by choosing a particular course of action.

R5 Time Value of Money

- **Decompose required rate of return:**
  - Nominal risk-free rate = real risk-free rate + expected inflation rate
  - Required interest rate on a security

    = nominal risk-free rate + default risk premium + liquidity risk premium + maturity risk premium
- 考察方法：
  - Real risk-free rate和nominal risk-free rate的关系

R5 Example: Time Value of Money

- Now, the nominal risk-free rate decreases. Keep the credit risk, liquidity risk and maturity risk constant, if the inflation rate increases, the real risk-free rate will be:

  A. Decrease

  B. No change

  C. Increase

- Correct answer: A

R5 Time Value of Money

EAR calculation:

$$\boxed{\text{EAR}=(1+\text{periodic rate})^{m}-1} \longleftrightarrow 1+EAR=\left(1+\frac{r}{m}\right)^{m}=e^{r}$$

- 那么如果是**semi, m=2;** 如果是**quarterly, m=4**
- 如果是连续复利，公式则变为**EAR = e $^{\text{annual int}}$-1**

考察方法：

- 计算——算**EAR**，或者是算计息次数
- 定性（EAR和计息次数有关）
  - The greater the compounding frequency,
    - ✓ the greater the EAR will be in comparison to the stated rate
    - ✓ the greater the difference between EAR and the stated rate

R5 Example: Time Value of Money

- A money manager has $1,000,000 to invest for one year. She has identified two alternative one-year certificates of deposit (CD) shown below:

	Compounding frequency	Annual interest rate
CD1 CD2	Quarterly Continuously	4.00% 4.95%

- Which CD has the highest effective annual rate (EAR) and how much interest will it earn?

	Highest EAR	Interest earned
A.	CD1	$41,902
B.	CD1	$40,604
C.	CD2	$50,700

R5 Example: Time Value of Money

- John plans to invest \$2,500 in an account that will earn 8% per year with quarterly compounding. How much will be in the account at the end of two years?
- **Answer:**

There are eight quarterly compounding periods in two years, and the effective quarterly rate is 8/4=2%.

The account will grow to $2{,}500(1.02)^8$=\$2,929.15.

Alternatively, since the EAR is 1.02^4-1=0.082432, we can grow the \$2,500 at 8.2432% for two years to get $2{,}500(1.082432)^2$=\$2,929.15, which is the same result.

R5 Time Value of Money

- Future value (FV): Amount to which investment grows after one or more compounding periods.
- Present value (PV): Current value of some future cash flow
- If interests are compounded m times per year, and invest 1 year:

  $$FV=PV(1+r/m)^{m}$$

- If interests are compounded m times per year, and invest n years:

  $$FV=PV(1+r/m)^{mn}$$

  Where: m is the compounding frequency;

  r is the nominal/quoted annual interest rate.

- When we calculate the future value of continuously compounding, the formula is:

$$FV=PV\lim_{m\to\infty}(1+\frac{r}{m})^{nm}=PVe^{nr}$$

R5 Time Value of Money

What's annuities?

--- is a stream of <u>equal cash flows</u> that occurs at <u>equal intervals</u> over a given period

- 内容：
  - N = number of periods
  - I/Y = interest rate per period
  - PV = present value
  - PMT = amount of each periodic payment
  - FV= future value
- 考察方法：计算——N, I/Y, PMT, FV, PV中任意给定四个，求另外一个

R5 Time Value of Money

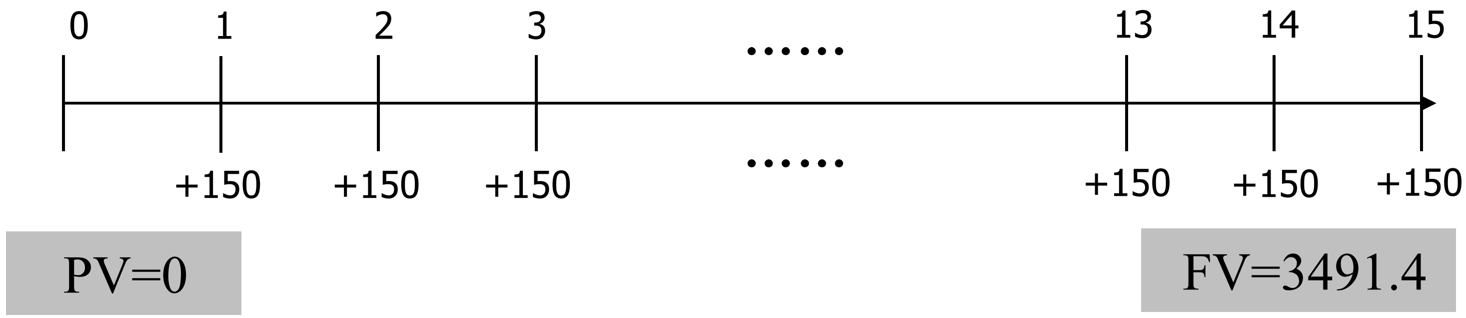
An example of **ordinary annuities（后付年金）:**

Example 1: What's the FV of an ordinary annuity that pays 150 per year <u>at the end of</u> each of the next15 years, given the discount rate is 6%

Solutions: enter relevant data for calculate.

N=15, I/Y=6, PMT=-150, PV=0, CPT→FV=3491.4

Notice: if we were given that FV= 3491.4, N=15, I/Y=6, PMT=-150, we also could calculate PV.



R5 Time Value of Money

- **About an annuity due（先付年金）**
  - **Definition:** an annuity where the annuity payments occur at the beginning of each compounding period.
  - **Calculation:**
    - ✓**Measure 1:** put the calculator in the BGN mode and input relevant data.
    - ✓**Measure 2:** treat as an ordinary annuity and simply multiple the resulting PV by (1+I/Y)

R5 Example: Time Value of Money

- What is the present value of four \$100 end-of-year payments if the first payment is to be received three years from today and the appropriate rate of return is 9%?

- **Answer:**
- Step 1: Find the present value of the annuity as of the end of year 2 (PV_2)

  N=4; I/Y=9; PMT=-100; FV=0; CPT: $PV=PV_2=\$323.97$
- Step 2: Find the present value of PV_2

  N=2; I/Y=9; PMT=0; FV=-323.97; CPT: $PV=PV_0=\$272.68$

R5 Example: Time Value of Money

- Construct an amortization schedule to show the interest and principal components of the end-of-year payments for a 10%, 5-year, $10,000 loan.
- **Answer:**

The amount of the loan payments: N=5; I/Y=10; PV=-$10,000; CPT: PMT=$2,637.97

Amortization Table

Period	Beginning Balance	Payment	Interest Component (1)	Principal Component (2)	Ending Balance (3)
1	$10,000.00	$2,637.97	$1,000.00	$1,637.97	$8,362.03
2	8,362.03	2,637.97	836.20	1,801.77	6,560.26
3	6,560.26	2,637.97	656.03	1,981.94	4,578.32
4	4,578.32	2,637.97	457.83	2,180.14	2,398.18
5	2,398.18	2,638.00*	239.82	2,398.18	0.00

* There is usually a slight amount of rounding error that must be recognized in the final period. The extra $0.03 associated with payment five reflects an adjustment for the rounding error and forces the ending balance to zero.

R5 Example: Time Value of Money

- Suppose you borrowed \$10,000 at 10% interest to be paid semiannually over ten years. Calculate the amount of the outstanding balance for the loan after the second payment is made.

- **Answer:**
- First, the amount of the payment must be determined by entering the relevant information and computing the payment.

  PV=-\$10,000; I/Y=10/2=5; N=10×2=20; CPT: PMT=\$802.43
- The principal and interest component of the second payment can be determined using the following process:

  Payment 1: Interest = (\$10,000)(0.05) = \$500

  Principal = \$802.43-\$500 = \$302.43

  Payment 2: Interest = (\$10,000-\$302.43)(0.05) = \$484.88

  Principal = \$802.43-\$484.88=\$317.55

  Remaining balance = \$10,000-\$302.43-\$317.55 = \$9,380.02

R5 Example: Time Value of Money

- Suppose you must make five annual $1,000 payments, the first one starting at the beginning of Year 4 (end of Year 3). To accumulate the money to make these payments, you want to make three equal payments into an investment account, the first to be made one year from today. Assuming a 10% rate of return, what is the amount of these three payments?
- **Answer:**
- The first step in this type of problem is to determine the amount of money that must be available at the beginning of Year 4 (t=3) in order to satisfy the payment requirements.

  BGN mode: N=5; I/Y=10; PMT=-1,000; CPT: PV=$4,169.87
- Determine the amount of the three payments.

  END mode: N=3; I/Y=10; PV=0; FV=-4,169.87; CPT: PMT=$1,259.78

R5 Example: Time Value of Money

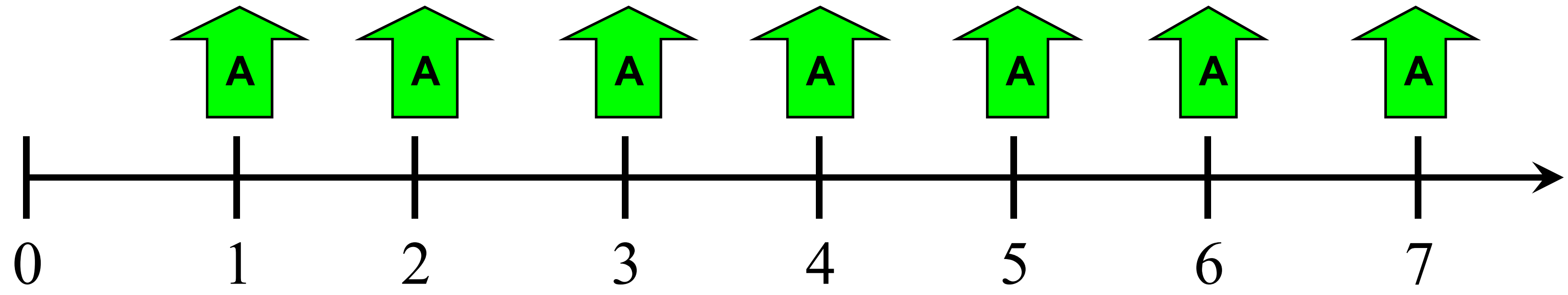
- Sales at Acme, Inc., for the last five years (in million) have been $4.5, $5.7, $5.3, $6.9, and $7.1. What is the compound annual growth rate of sales over the period?
- **Solution:**

The five years of sales represent four years of growth. Mathematically, the compound annual growth rate of sales is (7.1/4.5)1/4-1=12.1%. The interim sales figures do not enter into the 4-year compound growth rate.

Calculator solution: FV=7.1; PV=-4.5; N=4; CPT: I/Y=12.08%

R5 Time Value of Money

- **About** **perpetuity**
  - **Definition:** A perpetuity is a financial instruments that pays a fixed amount of money at set intervals over an **infinite** period of time.



$PV_1 = A/(1+r)$

$PV_2 = A/(1+r)^2$

$PV_3 = A/(1+r)^3$

$PV_4 = A/(1+r)^4$

etc.

etc.

$$PV = \frac{A}{1+r} + \frac{A}{(1+r)^2} + \frac{A}{(1+r)^3} + \cdots \qquad (1)$$

$$(1+r)PV = A + \frac{A}{1+r} + \frac{A}{(1+r)^2} + \cdots \qquad (2)$$

$$(2)-(1) \qquad r \times PV = A \Rightarrow PV = \frac{A}{r}$$

R5 Example: Time Value of Money

- In the same market, there are three kinds of investments, given a 5 percent discount rate for them.

  Investment 1: Perpetuity, \$100 per year starting in Year1 (first payment at t=1)

  Investment 2: Perpetuity, \$100 per year starting in Year5 (first payment at t=5)

  Investment 3: four-year ordinary annuity of \$100 per year starting in Year1

  The present value of the investment 3 is closest to:

- **Solution:**

  If we subtract Perpetuity 2 from Perpetuity 1, we are left with an ordinary annuity of \$100 per period for four years (payments at t=1,2,3,4). Subtracting the present value of Perpetuity 2 from that of Perpetuity 1, we arrive at the present value of the four-year ordinary annuity:

  PV_0(Perpetuity 1) = \$100/0.05 = \$2,000

  PV_4(Perpetuity 2) = \$100/0.05 = \$2,000

  PV_0(Perpetuity 2) = \$2,000/$(1.05)^4$ = \$1,645.40

  PV_0(Annuity) = PV_0(Perpetuity 1) – PV_0(Perpetuity 2) = \$2,000-\$1,645.40 = \$354.60

R6 Discounted Cash Flow Applications

- Discounted Cash Flow Applications
  1. NPV & IRR
  2. 计算HPY，EAY等收益率，以及它们相互之间的转化
  3. Money-weighted return & Time-weighted return

R6 Discounted Cash Flow Applications

$$NPV = CF_0 + \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \dots + \frac{CF_N}{(1+r)^N} = \sum_{t=0}^{N} \frac{CF_t}{(1+r)^t}$$

$$NPV = 0 = CF_0 + \frac{CF_1}{(1+IRR)^1} + \frac{CF_2}{(1+IRR)^2} + \dots + \frac{CF_N}{(1+IRR)^N} = \sum_{t=0}^{N} \frac{CF_t}{(1+IRR)^t}$$

IRR（Internal Rate of Return）

- When NPV= 0, the discount rate.
- IRR method assumes the project's cash flows will be reinvested at the IRR.
- Multiple solutions Problem of the IRR calculation (# sign changes)

R6 Discounted Cash Flow Applications

➢Project Decision Rule

●Single project Case

- ✓ NPV method: Accept it if NPV>0
- ✓ IRR method: Accept it if IRR>r (required rate of return)

●Two Projects Case

Independent Projects

- ✓ Similar to Single projects case

Mutually Exclusive Projects

- ✓ NPV method: Choose the one with higher NPV
- ✓ IRR method: Choose the one with higher IRR
- ✓ NPV and IRR methods may conflict with each other

R6 Example: Discounted Cash Flow Applications

1. Calabash Crab House is considering an investment in kitchen-upgrade projects with the following cash flows:

	Project A	*Project B*
Initial Year	-$10,000	-$9,000
Year 1	2,000	200
Year 2	5,000	-2,000
Year 3	8,000	11,000
Year 4	8,000	15,000

- Assuming Calabash has a 12.5 percent cost of capital, which of the following investment decisions has the *least* justification? Accept:
  - A. Project B because the net present value (NPV) is higher than that of Project A.
  - B. Project A because the IRR is higher than the cost of capital.
  - C. Project A because the internal rate of return (IRR) is higher than that of Project B.
- **Correct answer: C**

R6 Example: Discounted Cash Flow Applications

2. Which of the following statements least accurately describes the IRR and NPV methods?
   A. The discount rate that gives an investment an NPV of zero is the investment's IRR.
   B. If the NPV and IRR methods give conflicting decisions for mutually exclusive projects, the IRR decision should be used to select the project.
   C. The NPV method assumes that a project's cash flows will be reinvested at the cost of capital, while the IRR method assumes they will be reinvested at the IRR.

- **Answer: B**

3. Which of the following statements least accurately describes the IRR and NPV methods?
   A. A project's IRR can be positive even if the NPV is negative.
   B. A project with an IRR equal to the cost of capital will have an NPV of zero.
   C. A project's NPV may be positive even if the IRR is less than the cost of capital.

- **Answer: C**

R6 Example: Discounted Cash Flow Applications

4. Which of the following statements least accurately describes the IRR and NPV methods?
   A. The NPV tells how much the value of the firm has increased if you accept the project.
   B. When evaluating independent projects, the IRR and NPV methods always yield the same accept/reject decisions.
   C. When selecting between mutually exclusive projects, the project with the highest NPV should be accepted regardless of the sign of the NPV calculation.

➢ **Answer: C**

R6 Discounted Cash Flow Applications

HPR

- **Define:** the holding period return is simply the percentage change in the value of an investment over the period it is hold.

- **Calculate:**

$$\boxed{HPR = \frac{P_1 - P_0 + CF_1}{P_0}}$$

R6 Discounted Cash Flow Applications

$$r_{BD} = \frac{(F - P_0)}{F} \times \frac{360}{t}$$

$$HPY = \frac{P_1 - P_0 + CF_1}{P_0}$$

$$EAY = (1 + HPY)^{365/t} - 1$$

$$\mathrm{HPY} = \frac{\mathrm{r_{BD}}(\frac{\mathrm{t}}{360})}{1 - \mathrm{r_{BD}}(\frac{\mathrm{t}}{360})}$$

$$r_{MM} = HPY \times \frac{360}{t} = \frac{360 \cdot r_{BD}}{360 - t \cdot r_{BD}} = \frac{r_{BD}}{1 - t \cdot r_{BD} / 360}$$

$$(1 + \frac{\mathrm{BEY}}{2})^2 = 1 + \mathrm{EAR}$$

R6 Discounted Cash Flow Applications

- The **HPY** is the actual return an investor will receive if the money market instrument is held until maturity.
- The **EAY** is the annualized HPY on the basis of a 365-day year and incorporates the effects of compounding.
- The $\mathbf{r_{MM}}$ is the annualized yield that is based on price and a 360-day year and dose not account for the effects of compounding – it assumes simple interest.

R6 Example: Discounted Cash Flow Applications

- Jane Peebles purchased a T-bill that matures in 200 days for $97,500. The face value of the bill is $100,000. What is the money market yield on the bill?

  A. 4.500%.

  B. 4.615%.

  C. 4.756%.

- **Correct answer: B**

R6 Example: Discounted Cash Flow Applications

- A 175-day T-bill has an effective annual yield of 3.80%. Its bank discount yield is closest to:

  A. 1.80%

  B. 3.65%

  C. 3.71%

- **Answer: B**

R6 Discounted Cash Flow Applications

Money-weighted and time-weighted Rate of Return

- time-weighted return掌握概念及公式：
  - 概念：Time-weighted rate of return measures compound growth.
  - 步骤及公式：Firstly, compute the HPR; then, compute (1+HPR) for each subperiod to obtain a total return for the entire measurement period [eg. (1+HPR1) * (1+HPR2)…(1+HPRn)].
- money-weighted return掌握概念及公式：
  - 概念：the IRR based on the cash flows related to the investment
  - 步骤及公式：Firstly, determine the timing of each cash flow; then, using the calculation to compute IRR, or using geometric mean.
- 考察方法：计算；注意计算time-weighted return时，如果不是年度的HPR不用开方

R6 Example: Discounted Cash Flow Applications

- Assume an investor buys a share of stock for $100 at t = 0 and at the end of the next year (t = 1), she buys an additional share for $120. At the end of Year 2, the investor sells both shares for $130 each. At the end of each year in the holding period, the stock paid a $2.00 per share dividend.
  - What is the money-weighted rate of return?
  - What is the annual time-weighted rate of return?
- Solution:
  - Money-weighted rate of return (IRR) =13.86%
  - Time-weighted rate of return (geometric mean return) = 15.84%

R6 Example: Discounted Cash Flow Applications

- Would a client making additions or withdrawals of funds most likely affect their portfolio's:

	Time-weighted return?	Money-weighted return?
A.	No	No
B.	No	Yes
C.	Yes	No

- Correct answer: B
- **Solution**
  - The time-weighted return is not affected by cash withdrawals or addition to the portfolio, the money-weighted return measure would be affected by client additions or withdrawals, if a client adds funds at a favorable time the money-weighted return will be elevated.

R7 Statistical Concepts and Market Return

- Statistical concepts

1. Types of measurement scales
2. Measures of central tendency
3. Quantile
4. MAD和Var计算以及比较
5. Chebyshev's inequality
6. CV & Sharp ratio
7. Skewness & Kurtosis

R7 Statistical Concepts and Market Return

- Descriptive statistics
  - Summarize the important characteristics of large data sets.
- Inferential statistics
  - Make forecasts, estimates, or judgments about a large set of data on the basis of the statistical characteristics of a smaller set (a sample)

R7 Statistical Concepts and Market Return

Types of measurement scales:

- **Nominal scales**
  - distinguishing two different things, no order, only has mode
  - example: assigning the number 1 to a municipal bond fund, the number 2 to a corporate bond fund.
- **Ordinal scales (>, <)**
  - making things in order, but the difference are not meaningful
  - example: the ranking of 1,000 small cap growth stocks by performance may be done by assigning the number 1 to the 100 best performing stocks
- **Interval scales (>, <, +, -)**
  - subtract is meaningful
  - example: Temperature
- **Ratio scales (>, <, +, -, *, /)**
  - with original point
  - example: money, if you have zero dollars, you have no purchasing power, but if you have $4.00, you have twice as much purchasing power as a person with $2.00.

R7 Example: Statistical Concepts and Market Return

- An analyst gathered the price-earnings ratios (P/E) for the firms in the S&P 500 and then ranked the firms from highest to lowest P/E. She then assigned the number 1 to the group with the lowest P/E ratios, the number 2 to the group with the second lowest P/E ratios, and so on. The measurement scale used by the analyst is *best* described as:
  - A. Ratio.
  - B. Ordinal.
  - C. Interval.
- Correct answer: B

R7 Statistical Concepts and Market Return

- A measure used to describe a characteristic of a population is referred to as a **parameter**.
- In the same manner that a parameter may be used to describe a characteristic of a population, a **sample statistic** is used to measure a characteristic of a sample.

R7 Statistical Concepts and Market Return

- **Relative frequency**
  - The relative frequency is calculated by dividing the absolute frequency of each turn interval by the total number of observations.
- **Frequency Distribution**
  - A frequency distribution is a tabular presentation of statistical data that aids the analysis of large data sets.
- **Cumulative frequency/Cumulative Relative Frequency**
  - Could be calculated by summing the absolute or relative frequencies starting at the lowest interval and progressing through the highest.

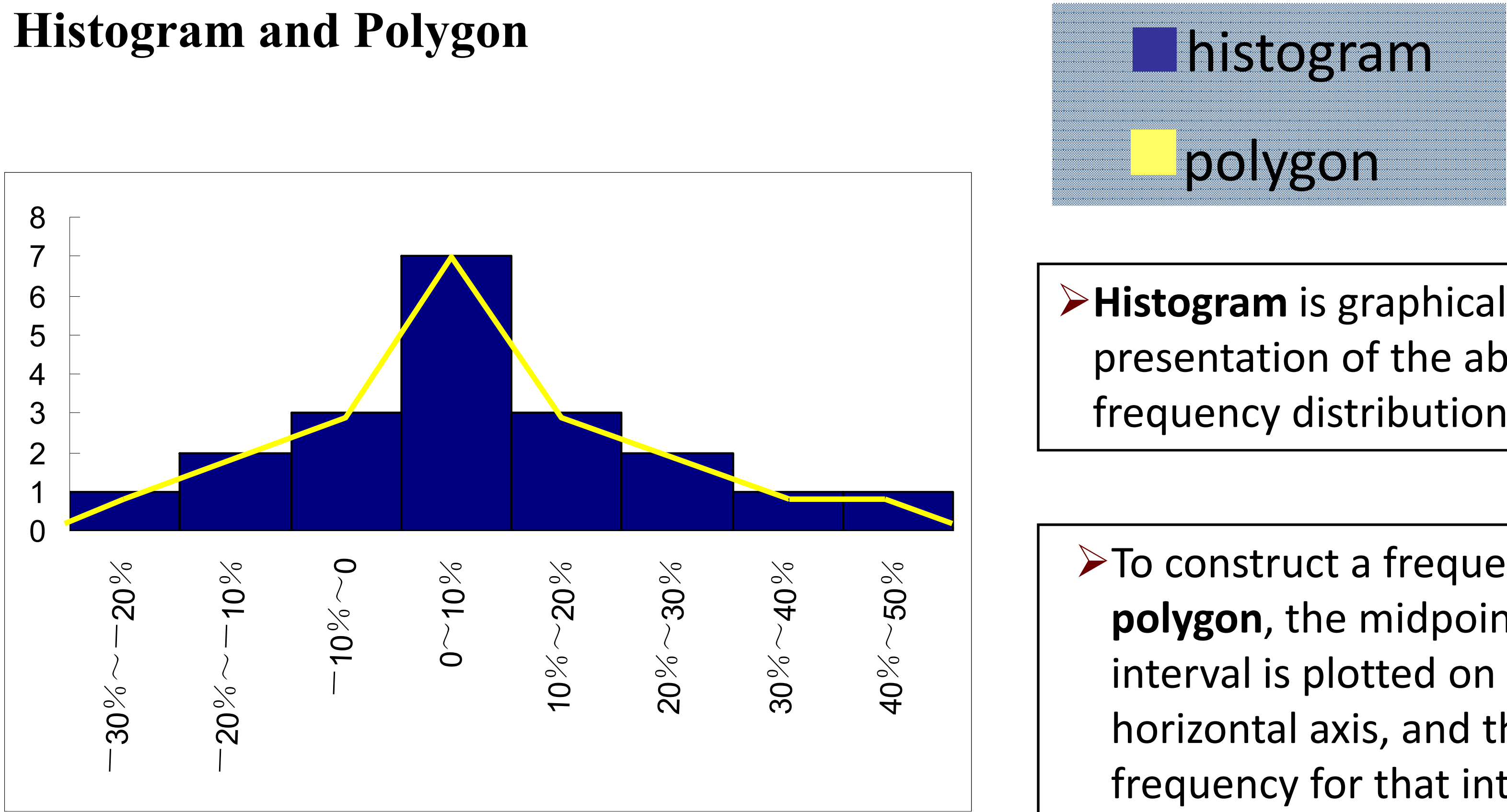
R7 Statistical Concepts and Market Return

Frequency distribution

Interval Relative	Absolute Frequency	Relative Frequency	Cumulative Absolute Frequency	Cumulative Frequency
-10 - -5	*3*	*0.97%*	*3*	*0.97%*
-5 – 0	*35*	*11.29%*	*38*	*12.26%*
0 – 5	*176*	*56.77%*	*214*	*69.03%*
5 – 10	*74*	*23.87%*	*288*	*92.90%*
10 - 15	*22*	*7.10%*	*310*	*100%*
Total	***310***	***100%***		

R7 Statistical Concepts and Market Return

Histogram and Polygon



- **Histogram** is graphical presentation of the absolute frequency distribution

- To construct a frequency **polygon**, the midpoint of each interval is plotted on the horizontal axis, and the absolute frequency for that interval is plotted on the vertical axis.

R7 Example: Statistical Concepts and Market Return

- An analyst gathered the following information about the price-earning (P/E) ratios for the common stocks held in a foundation's portfolio:

Interval	P/E range	Frequency
I	8.00-16.00	24
II	16.00-24.00	48
III	24.00-30.00	22
IV	30.00-38.00	16

- The relative frequency and the cumulative relative frequency, respectively, for interval III are closest to:

	Relative frequency	Cumulative relative frequency
A.	20%	85%
B.	22%	36%
C.	22%	85%

- **Solution: A**

R7 Statistical Concepts and Market Return

The arithmetic mean:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{n}$$

The weighted mean:

$$\overline{X_W} = \sum_{i=1}^{n} w_i X_i = (w_1 X_1 + w_2 X_2 + \cdots + w_n X_n)$$

The geometric mean:

$$G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N} = \left(\prod_{i=1}^{N} X_i\right)^{1/N}$$

The harmonic mean:

$$\overline{X_H} = \frac{n}{\sum_{i=1}^{n} (1/X_i)}$$

harmonic mean<= geometric mean<=arithmetic mean

R7 Example: Statistical Concepts and Market Return

- Which is the most accurate?

	Harmonic mean	Arithmetic mean	Geometric mean
A.	13	15	18
B.	15	15	18
C.	13	18	15

- Correct answer: C

R7 Example: Statistical Concepts and Market Return

- An analyst obtains the following annual rates of return for a mutual fund:

Year	Return (%)
2008	14
2009	-10
2010	-2

1. The fund's holding period return over the three-year period is closest to:
   A. 0.18%
   B. 0.55%
   C. 0.67%

- **Answer: B**

2. The fund's annual holding period return is closest to:
   A. 0.18%
   B. 0.55%
   C. 0.67%

- **Answer: A**

R7 Example: Statistical Concepts and Market Return

- A hypothetical investment in a single stock initially costs \$100. one year later, the stock is trading at \$200. At the end of the second year, the stock price falls back to the original purchase price of \$100. No dividend are paid during the two-year period. Calculate the arithmetic and geometric mean annual returns.
- **Solution:**

Return in Year1 = 200/100-1 = 100%

Return in Year2 = 100/200-1 = -50%

Arithmetic mean = (100%-50%)/2 =25%

Geometric mean = $(2.0\times 0.5)^{1/2} - 1 = 0\%$

The geometric mean return of 0% accurately reflects that the ending value of the investment in Year2 equals the starting value in Year1. The compound rate of return on the investment is 0%. The arithmetic mean return reflects the average of the one-year returns.

R7 Statistical Concepts and Market Return

- The use of arithmetic mean and geometric mean when determining investment returns
  - The arithmetic mean is the statistically best estimator of the next year's returns given only the three years of return outcomes.
  - Since past annual returns are compounded each period, the geometric mean of past annual returns is the appropriate measure of past performance.

R7 Statistical Concepts and Market Return

- Quantiles
  - Quartile /Quintile/Deciles/Percentile
    - ✓The third quartile: 75%, or three-fourths of the observations fall below that value.
  - Calculation $L_y = (n+1)y/100$, Ly is the position.

Example:

Observers: **8 10 12 13 15 17 17 18 19 23 24**

N=11， Ly=(11+1)*75%=9,i.e. the 9th number is 75%

The third quartiles = 19

R7 Statistical Concepts and Market Return

- **Absolute dispersion**: is the amount of variability present without comparison to any reference point or benchmark.

Range = maximum value – minimum value

$$MAD = \frac{\sum_{i=1}^{N} \left| X_i - \bar{X} \right|}{n}$$

$$\text{For population: } \sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

$$\text{For sample: } s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

$$\text{Semivariance} = \frac{\sum_{\text{for all } X_i \le \bar{X}} (X_i - \bar{X})^2}{n-1}$$

$$\text{Target Semivariance} = \frac{\sum_{\text{for all } X_i \le B} (X_i - B)^2}{n-1}$$

R7 Example: Statistical Concepts and Market Return

- An analyst gathered the following annual return information about a portfolio since its inception on 1 January 2003:

Year	Portfolio return
2003	8.6%
2004	11.2%
2005	12.9%
2006	15.1%
2007	-9.4%

- The portfolio's mean absolute deviation and variance of annual returns, respectively, for the five-year period are closest to:

	Mean absolute deviation	variance
A.	6.83%	77.5
B.	6.83%	96.8
C.	7.68%	77.5

- **Answer: A**

R7 Statistical Concepts and Market Return

- For any set of observations (samples or population), the proportion of the values that lie within k standard deviations of the mean is at least $1 – 1/k^2$, where k is any constant greater than 1.
- 对任何一组观测值，个体落在均值周围k个标准差之内的概率不小于$1-1/k^2$,对任意$k>1$。
- This relationship applies regardless of the shape of the distribution

R7 Example: Statistical Concepts and Market Return

- Assume a sample of beer prices is negatively skewed. Approximately what percentage of the distribution lies within plus or minus 2.40 standard deviations of the mean?

  A. 82.6%

  B. 58.3%

  C. 17.36%

- Correct answer: A

R7 Example: Statistical Concepts and Market Return

- The arithmetic mean monthly return and standard deviation of monthly returns on the S&P 500 were 0.97 percent and 5.65 percent, respectively, during the 1926-2002 period, totaling 924 monthly observations. Using this information, address the following:

1. Calculate the endpoints of the interval that must contain at least 75 percent of monthly returns according to Chebyshev's inequality.
2. What are the minimum and maximum number of observations that must lie in the interval computed in Part 1, according to Chebyshev's inequality?

- **Solution to 1:**

  $1-1/k^2=75\% \rightarrow k=2$

  $0.97\% \pm 2(5.65\%)=0.97\% \pm 11.30\%$

  Lower endpoint of the interval: 0.97%-11.30%=-10.33%

  Upper endpoint of the interval: 0.97%+11.30%=12.27%

- **Solution to 2:**

  For a sample size of 924, at least 0.75(924)=693 observations must lie in the interval from -10.33% to 12.27% that we computed in Part 1.

R7 Statistical Concepts and Market Return

- **Coefficient of variation** measures the amount of dispersion in a distribution relative to the distribution's mean. (relative dispersion)

$$\boxed{CV=\frac{s_x}{\overline{\overline{X}}}\times 100\%}$$

- **The sharp ratio** measures excess return per unit of risk.

$$\boxed{\text{Sharp ratio}=\frac{R_P - R_f}{\sigma_P}}$$

R7 Example: Statistical Concepts and Market Return

- An analyst gathered the following information about a portfolio's performance over the past ten years:

Mean annual return	12.8%
Mean excess return	7.4%
Standard deviation of annual returns	15.7%
Portfolio beta	1.2

- The coefficient of variation and Sharpe measure, respectively, for the portfolio are *closest* to:

	Coefficient of variation	Sharpe measure
A	0.82	0.39
B	0.82	0.47
C	1.23	0.47

- Correct answer: C

R7 Example: Statistical Concepts and Market Return

- The scale-free measure of relative dispersion that is useful in making direct comparisons among different asset classes is the:

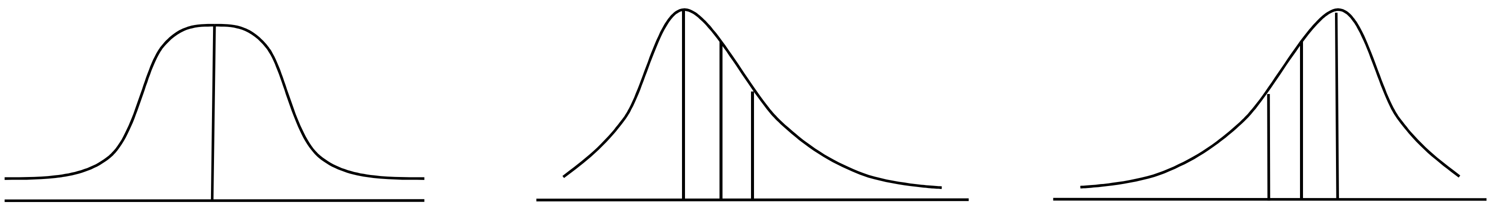
  A. Range.

  B. Variation.

  C. Coefficient of variation.

- Correct answer: C

R7 Statistical Concepts and Market Return



Mean=Median=Mode | Mode<Median<Mean | Mean<Median<Mode

Symmetrical | **Positive (right) skew** | **Negative (left) skew**

- **Positive skewed:** Mode<median<mean, having a right fat tail
  - A return distribution with positive skew has frequent small losses and a few extreme gains
- **Negative skewed:** Mode>media>mean, having a left fat tail
  - A return distribution with negative skew has frequent small gains and a few extreme losses.
- Investors should be attracted by a ***positive skew*** because the mean return falls above the median.
- **Sample skewness:** $S_K = \left[\frac{n}{(n-1)(n-2)}\right]\frac{\sum_{i=1}^{n}(X_i-\overline{X})^3}{s^3} \approx \left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(X_i-\overline{X})^3}{s^3}$

考察方法:

- 根据描述的特点判断是Positively skewed还是Negative skewed
- 根据已知的偏度，选择都有哪些特点

R7 Example: Statistical Concepts and Market Return

- A distribution with mode 2.6, median 2.2, mean 2, the distribution can be described as:
  - A. long tail in the left and positively skewed.
  - B. long tail in the right and negatively skewed.
  - C. long tail in the left and negatively skewed.
- **Solution: C**

- Which of the following is most accurate regarding a distribution of returns that has a mean greater than its median?
  - A. It is positively skewed.
  - B. It is a symmetric distribution.
  - C. It is negatively skewed.
- **Solution: A**

R7 Example: Statistical Concepts and Market Return

- As analyst gathered the following information about the return distribution of four investment. Based only on the information above, a well-diversified investor would most likely prefer Portfolio:

Portfolio	Skewness	Sharp Ratio
1	Positive	0.6
2	Positive	0.8
3	Negative	0.6

A. 1

B. 2

C. 3

- Correct answer: B

R7 Statistical Concepts and Market Return

- **Leptokurtic vs. platykurtic**
  - It deals with whether or not a distribution is more or less “peaked” than a normal distribution
- **Excess kurtosis = sample kurtosis – 3**

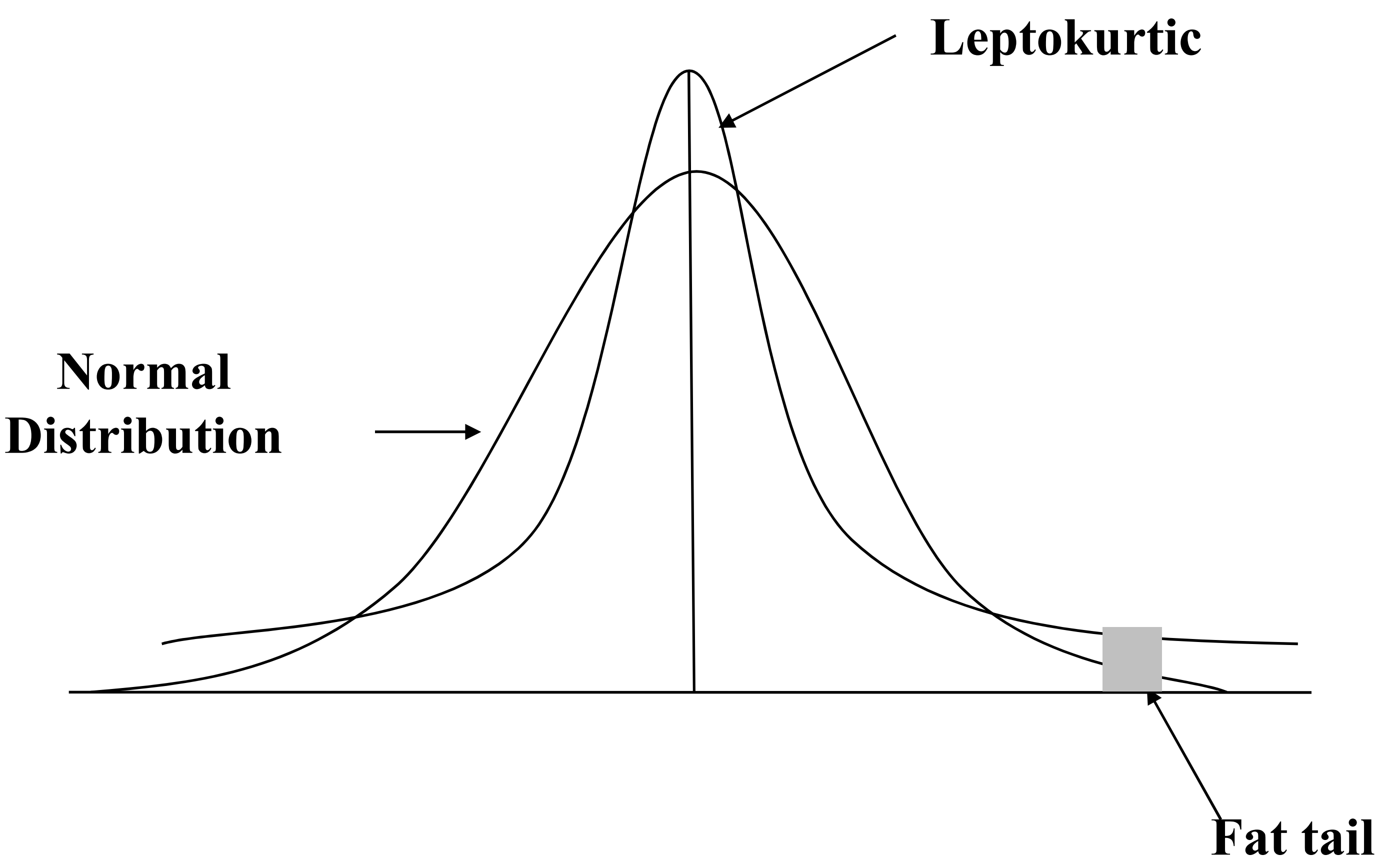
	leptokurtic	Normal distribution	platykurtic
Sample kurtosis	>3	=3	<3
Excess kurtosis	>0	=0	<0

- **Sample Kurtosis:**

$$K_E = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{s^4} \approx \frac{1}{n} \frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{s^4}$$

- 考察方法：
  - 根据描述的特点判断是leptokurtic还是platykurtic
  - 根据已知的峰度，选择都有哪些特点
  - 可能在考试中会和skew合并考核综合知识

R7 Statistical Concepts and Market Return



A leptokurtic return distribution has more frequent extremely large deviations from the mean than a normal distribution.

R7 Example: Statistical Concepts and Market Return

- An analyst gathered the following information about the return distribution for two portfolios during the same time period:

Portfolio	skewness	kurtosis
A	-1.3	2.2
B	0.5	3.5

- The analyst stated that the distribution for Portfolio A is more peaked than a normal distribution and that the distribution for Portfolio B has a long tail on the left side of the distribution. Is the analyst's statement correct with respect to:

	Porfolio A	Portfolio B
A.	No	No
B.	No	Yes
C.	Yes	No

- Solution: A

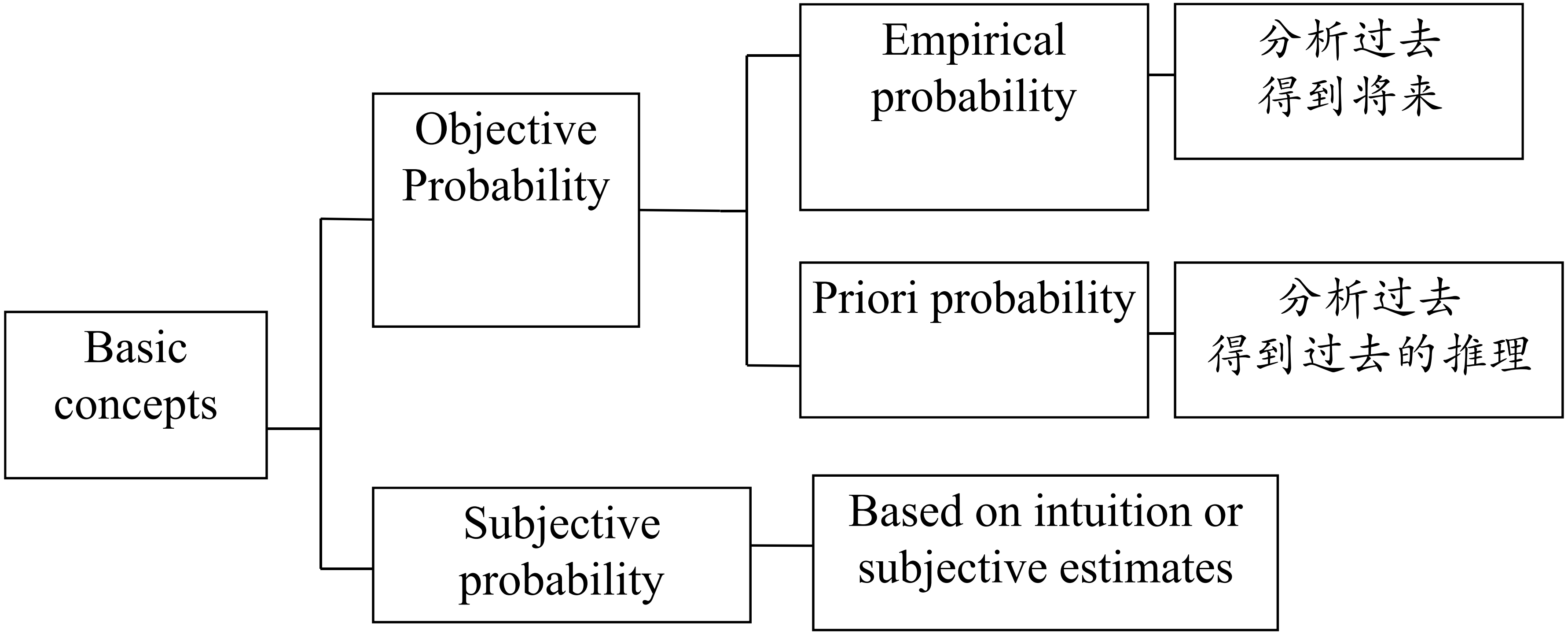
R8 Probability Concepts

- Probability concepts
  - Two defining properties of probability
  - Empirical, subjective, and priori probabilities
  - Odds for or against
  - Multiplication rule and addition rule
  - Dependent and independent events
  - Covariance & correlation
  - Expected value, variance, and standard deviation of a random variable and of returns on a portfolio
  - Bayes' formula

R8 Probability Concepts

- **Basic Concepts**
  - **Random variable** is uncertain quantity/number.
  - **Outcome is** an observed value of a random variable.
  - **Event**
    - ✓ Mutually exclusive events—can not both happen at the same time.
    - ✓ Exhaustive events—include all possible outcomes.
- **Two Defining Properties of Probability**
  - $0 \leq P(E) \leq 1$
  - $P(E1) + P(E2) + \ldots\ldots + P(En) = 1$

R8 Probability Concepts



100% Contribution Breeds Professionalism

R8 Probability Concepts

- **Empirical probability 经验概率**
  - eg. Historically, the Dow Jones Industrial Average has closed higher than the previous close two out of every three trading days. Therefore, the probability of the Dow going up tomorrow is two-thirds, or 66.7%.
- **Priori probability 先验概率**
  - eg. Yesterday, 24 of the 30 DJIA stocks increased in value. Thus, if 1 of 30 stocks is selected at random, there is an 80%(24/30) probability that its value increased yesterday
- **Subjective probability 主观概率**
  - will close higher tomorrow is 90%.

R8 Example: Probability Concepts

- An analyst adjusts the historical probability of default for high-yield bonds to reflect her perceptions of changes in the quality of high-yield bonds. The analyst is best characterized as obtaining a(n):

  A. A priori probability.

  B. Objective probability.

  C. Subjective probability.

- **Correct answer: C**

R8 Probability Concepts

- **Odds for an event**
  - P(E)/(1-P(E))
- **Odds against an event**
  - (1-P(E))/P(E)

Example:

- Last year, the average salary increase for Poultry Research Assistants was 2.5 percent. Of the 10,000 Poultry Research Assistants, 2,000 received raises in excess of this amount. The odds that a Poultry Research Assistant received a salary increase in excess of 2.5 percent are:

  A. 1 to 4.

  B. 2 to 10.

  C. 20%.

- Correct answer: A

R8 Probability Concepts

- Unconditional Probability (marginal probability): $P(A)$
- Conditional probability : $P(A|B)$

R8 Probability Concepts

- **Joint probability : P(AB)**
  - **Multiplication rule:**
    - ✓$P(AB)=P(A|B)\times P(B)= P(B|A)\times P(A)$
  - If A and B are mutually exclusive events, then:

    $P(AB)=P(A|B)=P(B|A)=0$
- **Probability that at least one of two events will occur:**
  - **Addition rule:**
    - ✓$P(A \text{ or } B)=P(A)+P(B)-P(AB)$
  - If A and B are mutually exclusive events, then:

    $P(A \text{ or } B)=P(A)+P(B)$

R8 Example: Probability Concepts

- The probability that two or more events will happen concurrently is best characterized as:

  A. Joint probability.

  B. Multiple probabilities.

  C. Concurrent probability.

- Correct answer: A

R8 Probability Concepts

- **The occurrence of A has <u>no influence</u> of on the occurrence of B**
  - $P(A|B)=P(A)$ or $P(B|A)=P(B)$
  - $P(AB)=P(A)\times P(B)$
  - $P(A \text{ or } B)=P(A)+P(B)-P(AB)$
- **Independence and Mutually Exclusive are quite different**
  - If exclusive, must not independence;
  - Cause exclusive means if A occur, B can not occur, A influents B.
    - ✓ $P(AB)=P(A)\times P(B)$

R8 Example: Probability Concepts

- A fundamental analyst studying 100 potential companies for inclusion in her stock portfolio uses the following three screening criteria:

Screening Criterion	Number of Companies meeting screen
Market-to-Book Ratio >4	20
Current Ratio >2	40
Return on Equity >10%	25

- Assuming that the screening criteria are independent, the probability that a given company will meet all three screening criteria is closest to:
  A. 2.0%.
  B. 8.5%.
  C. 20.0%
- Correct answer: A

R8 Example: Probability Concepts

- P (A) =0.5, P (B) =0.5, odd for concurrent A and B is 3/5, the relationship between A and B?

  A. dependent

  B. Independent

  C. Mutually exclusive

- Correct answer: A

- **Solution**

  - P(AB)=(3/5)/(1+3/5), P(A/B)=P(AB)/P(B)=3/4, P(A/B)不等于P(A)

R8 Probability Concepts

- For unconditional probability of event A,

$$P(A) = P(A|S_1)P(S_1) + P(A|S_2)P(S_2) + \ldots + P(A|S_N)P(S_N)$$

where the set of events $\{S_1, S_2, \ldots S_N\}$ is mutually exclusive and exhaustive.

- Expected value: $E(X) = \sum P(X_i)X_i$

$$\mathrm{E(X)} = \sum \mathrm{x_i} * \mathrm{P}(x_i) = x_1 * P(x_1) + x_2 * P(x_2) + \cdots + x_n * P(x_n)$$

$$\sigma = \sqrt{\sigma^2} \qquad \sigma^2 = \sum_{i=1}^{N} P_i (X_i - EX)^2$$

R8 Example: Probability Concepts

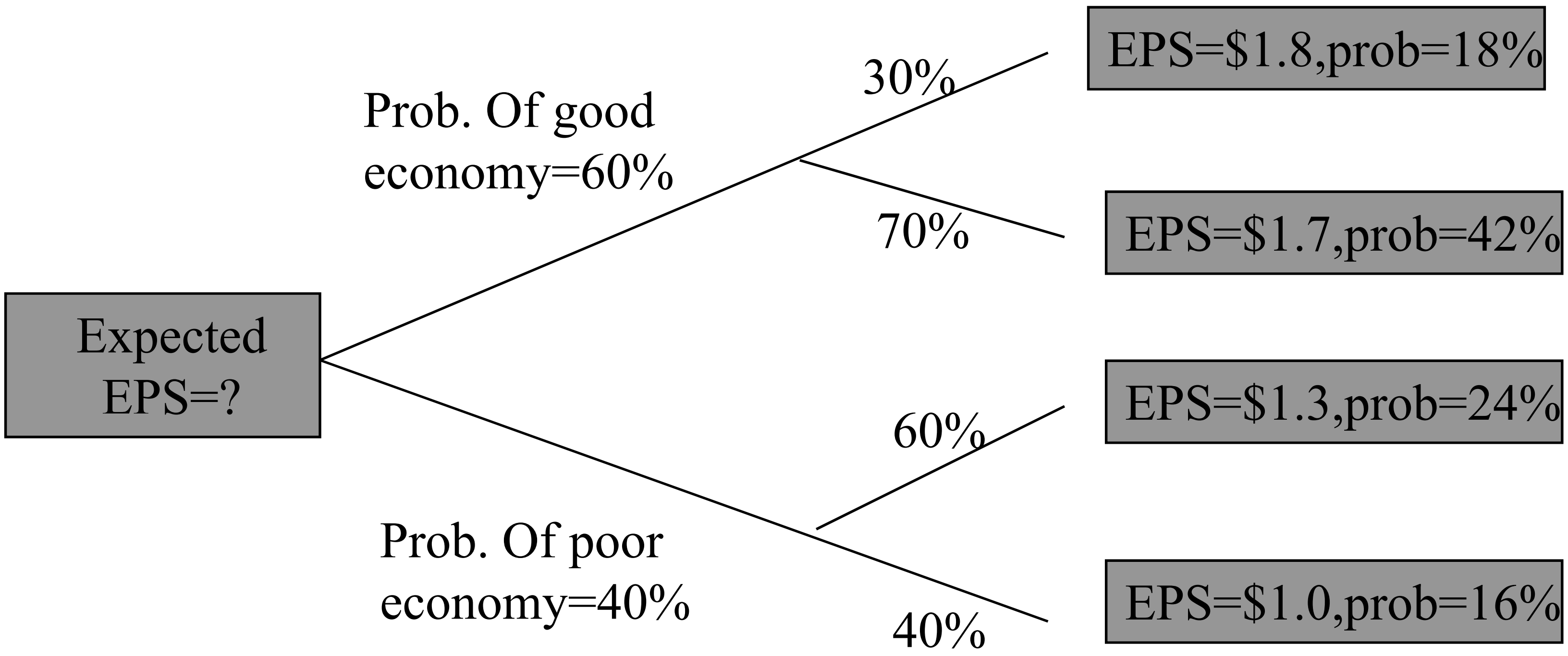
- An analyst gathered the following information: the probability of economy prosperity is 75%, the probability of economy recession is 25%. For a company, when the economy is prosperity, there is 10% of probability that its EPS is $2.0 and 90% of probability that the EPS is $4.0. However, when the economy is recession, there is 25% of probability that the EPS is $2.0 and 75% of probability that the EPS is $4.0. What is the variance of this company's EPS, when the economy is recession?
  - A. 3.55
  - B. 1.25
  - C. 0.75
- **Correct answer: C**

  When the economy recession:

  $E(EPS) = 25\% * 2 + 75\% *4 = 3.5$

  $Var(EPS) = 25\% * (2-3.5)^2 + 75\% * (4-3.5)^2 = 0.75$

R8 Probability Concepts



$$E(EPS) = 18\% \times 1.8 + 42\% \times 1.7 + 24\% \times 1.3 + 16\% \times 1.0 = 1.51$$

R8 Probability Concepts

- **Covariance:**
  - Covariance measures <u>how one random variable moves with another random variable</u>
  - The covariance of R_A with itself is equal to the variance of R_A
  - Covariance ranges from <u>negative infinity to positive infinity</u>

$$COV(X,X) = E[(X-E(X))(X-E(X))] = \sigma^2(X)$$

$$COV(X,Y) = E[(X-E(X))(Y-E(Y))]$$

- **Correlation:** $\rho_{XY} = \dfrac{COV(X,Y)}{\sqrt{Var(X)Var(Y)}}$
  - Correlation measures the **linear relationship** between two random variables
  - Correlation has no units, ranges from –1 to +1, standardization of covariance
  - Understand the difference between correlation and independence
  - If ρ=0, this indicates?

R8 Example: Probability Concepts

- The covariance of returns for two stocks:

  A. must have a value between -1.0 and +1.0

  B. must have a value equal to the weighted average of the standard deviations of the returns of the two stocks

  C. will be positive if the actual returns on both stocks are consistently below their expected returns at the same time

- **Correct answer: C**

R8 Example: Probability Concepts

- The joint probability of returns, for securities A and B, are as follows:

Joint Probability Function of Security A and Security B Returns (Entries are joint probabilities)		
	Return on security B=30%	Return on security B=20%
Return on security A=25%	0.60	0
Return on security A=20%	0	0.40

- The covariance of the returns between securities A and B is closest to:

A. $3(\%)^2$.

B. $12\ (\%)^2$.

C. $24\ (\%)^2$.

- **Correct answer: B**

R8 Example: Probability Concepts

- The correlation of returns between Stocks A and B is 0.50. The covariance between these two securities is 0.0043, and the standard deviation of the return of Stock B is 26%. The variance of return for Stock A is:
  - A. 0.0011
  - B. 0.0331
  - C. 0.2656
- **Answer: A**

- The correlation coefficient that indicates the weakest linear relationship between variables is:
  - A. -0.75
  - B. -0.22
  - C. 0.35
- **Answer: B**

R8 Probability Concepts

- Expected return, variance and standard deviation of a portfolio

$$E(r_{\mathrm{p}}) = \sum_{i=1}^{n} w_i E(R_i)$$

$$\sigma^2{}_p = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \operatorname{cov}(R_i, R_j)$$

R8 Example: Probability Concepts

- An individual wants to invest $100,000 and is considering the following stocks:

stock	Expected Return	Standard Deviation of Returns
A	12%	15%
B	16%	24%

- The expected correlation of returns for the two stocks is +0.5. If the investor invests $40,000 in Stock A and $60,000 in Stock B, the expected standard deviation of returns on the portfolio will be:

  A. equal to 20.4%

  B. less than 20.4%

  C. greater than 20.4% because the correlation coefficient is greater than zero

- Correct answer: B

R8 Example: Probability Concepts

- An analyst gathered the following information about a portfolio comprised of two assets:

Asset	Weight %	Expected Return E(R)	Expected Standard Deviation E(σ)
X	60	11%	5%
Y	40	7%	4%

- If the covariance of returns for the two assets equals 0.75, then the expected return and expected standard deviation of the portfolio are closest to:

	Expected Return	Expected Standard Deviation
A.	8.6%	4.3%
B.	8.6%	18.7%
C.	9.4%	4.3%

- **Answer: C**

R8 Example: Probability Concepts

- An fund manager has a portfolio of two mutual funds, A and B, 75 percent invested in A, as shown in the following table.

Covariance Matrix		
Fund	A	B
A	625	120
B	120	196

- The correlation between A and B, and the portfolio standard deviation of return is closest to:

	Correlation between A and B	Portfolio standard deviation of return
A.	0.18	40.80%
B.	0.34	20.22%
C.	0.12	18.00%

- **Answer: B**

R8 Example: Probability Concepts

- The correlation between assets in a two–asset portfolio increases during a market decline. If there is no change in the proportion of each asset held in the portfolio or the expected standard deviation of the individual assets, the volatility of the portfolio is most likely to:

  A. increase.

  B. decrease.

  C. remain the same.

- **Solution: A**

R8 Probability Concepts

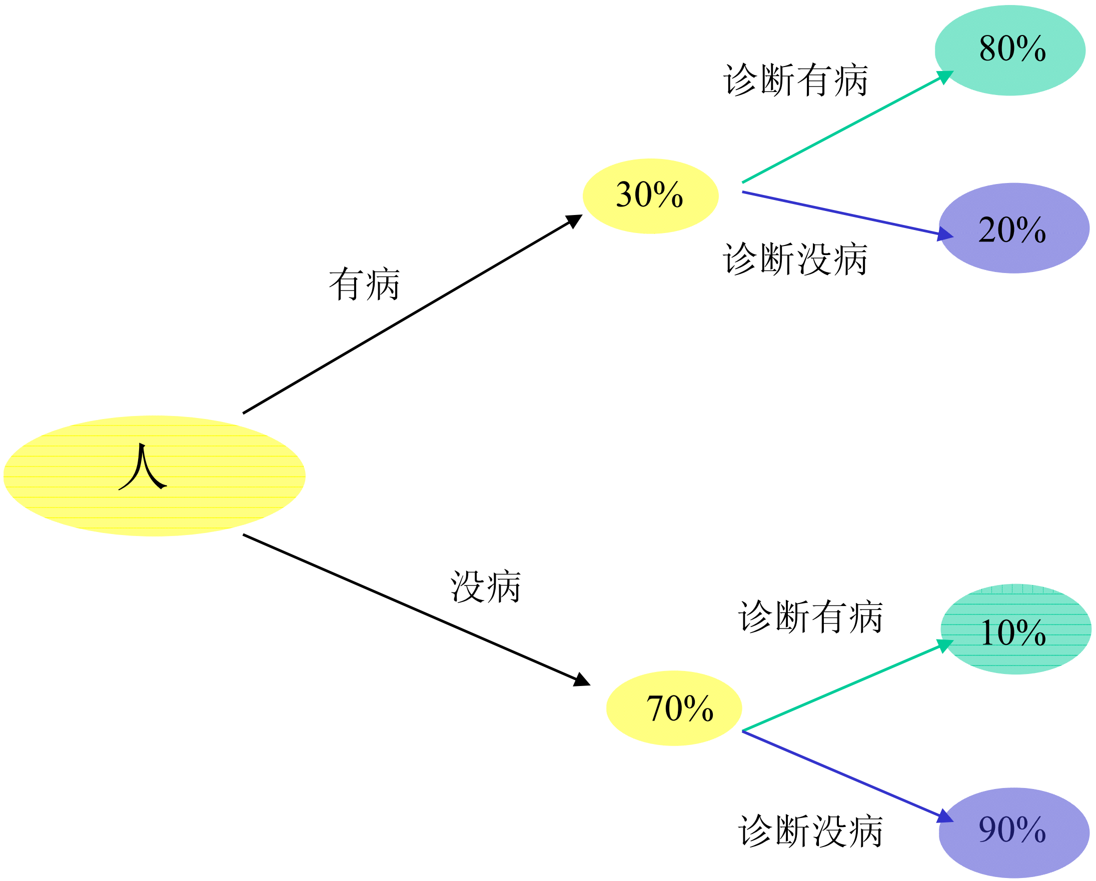
- **Bayes' Formula**
- **$P(AB)=P(A|B)\times P(B) =P(B|A)\times P(A)$**

$$P(A \mid B) = \frac{P(B \mid A)}{P(B)} * P(A)$$

- **$P(R)=P(R|S_1)\times P(S_1)+P(R|S_2)\times P(S_2)+\ldots+ P(R|S_n)\times P(S_n)$**

$$P(S_i \mid R) = \frac{P(R \mid S_i)P(S_i)}{P(R)}$$

R8 Probability Concepts



R8 Example: Probability Concepts

- An analyst has developed a ratio to identify companies expected to experience declining earnings per share (EPS). Research shows that 70 percent of firms experiencing a decline in EPS have a negative ratio, while only 20 percent of firms not experiencing a decline in EPS have a negative ratio. The analyst expects that 10 percent of all publicly traded companies will experience a decline in EPS next year. The analyst randomly selects a company and its ratio is negative. Based on Bayes' theorem, the posterior probability that the company will experience a decline in EPS next year is closest to:

  A. 14%

  B. 28%

  C. 30%

- Correct answer: B

R8 Probability Concepts

- Multiplication rule: $n_1 \times n_2 \times \ldots\ldots \times n_k$
- Factorial: $n!$
- Labeling (or Multinomial): $\dfrac{n!}{n_1! \times n_2! \times \ldots \times n_k!}$
- Combination: $${}_nC_r = \binom{n}{r} = \frac{n!}{(n-r)! \times r!}$$
- Permutation: $${}_nP_r = \frac{n!}{(n-r)!}$$

R8 Example: Probability Concepts

- How many ways are there to sell three stocks out of eight if the order of the sales is important?
  A. 56
  B. 336
  C. 6720
- **Solution: B**

- The probability that each stock in A stock market outperforms the Shanghai Composite index is 30%. For three stocks, What is the probability that at least two stocks outperformed the Shanghai Composite index?
  A. 21.6%
  B. 41.16%
  C. 78.4%
- **Solution: A**

R9 Common Probability Distributions

- Common Probability Distributions
  - Properties of discrete distribution and continuous distribution
  - Uniform random variable and a binomial random variable
  - The key properties of the normal distribution
  - Standardize a random variable
  - Confidence interval for a normally distributed random variable
  - Lognormal distribution
  - Safety-first ratio
  - Monte Carlo simulation

R9 Common Probability Distributions

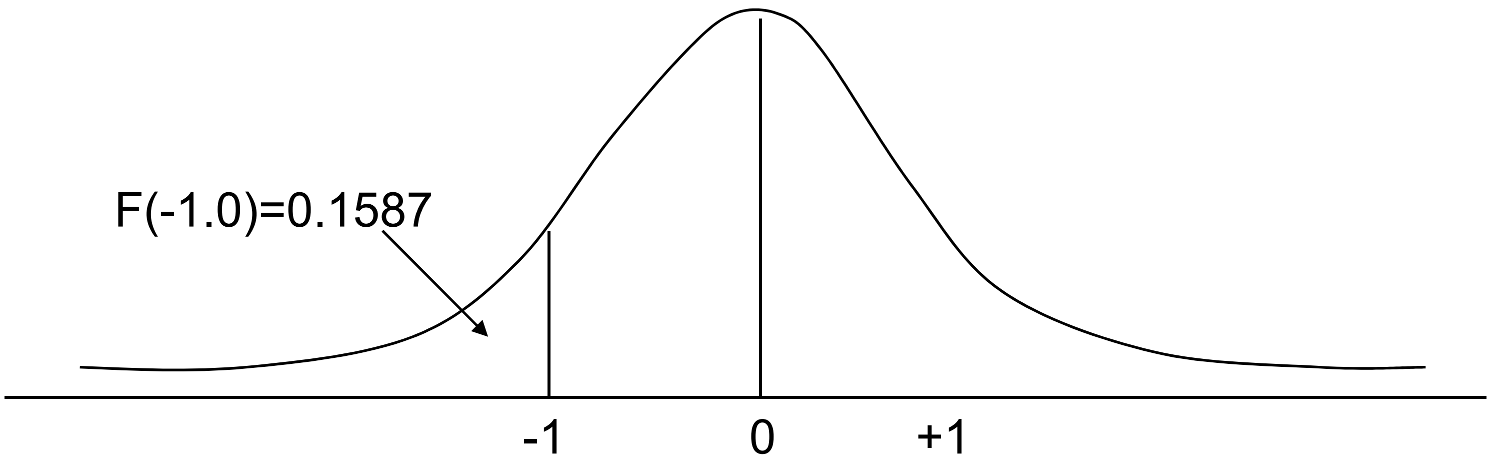
- **Probability Distribution**
  - Describe the probabilities of all the possible outcomes for a random variable.
- **Discrete and continuous random variables**
  - Discrete random variables: the number of possible outcomes can be counted, and for each possible outcome, there is a measurable and positive probability.
  - Continuous variables: the number of possible outcomes is infinite, even if lower and upper bounds exist.
    - ✓ P (x)=0 even though x can occur.
    - ✓ P (x1<X<x2)

R9 Common Probability Distributions

- **Probability function:** $p(x)=P(X=x)$
  - For discrete random variables
  - $0 \leq p(x) \leq 1$
  - $\Sigma p(x)=1$
- **Probability density function (p.d.f) :** $f(x)$
  - For continuous random variable commonly
- **Cumulative probability function (c.p.f) :** $F(x)$
  - $F(x)=P(X<=x)$

R9 Common Probability Distributions

➢ Probability density function



R9 Example: Common Probability Distributions

- Which of the following statements about probability distributions is **FALSE**?
  A. For a probability distribution for the number of days the air pollution is above a specified level, $p(x) = 0$ when x cannot occur, or $p(x) > 0$ when it can.
  B. For a probability distribution for the specific level of air pollution on a given day, $p(x) = 0$ even if x can occur.
  C. A cumulative distribution function gives the probability that a random variable takes a value equal to or greater than a given number.
- Correct answer: C
- **Solution**
  - A cumulative distribution function gives the probability that a random variable takes a value equal to or *less* than a given number: $P(X \leq x)$, or $F(X)$.

R9 Common Probability Distributions

- **Discrete uniform**
  - A discrete uniform random variable is one for which the probabilities for all possible outcomes for a discrete random variable are equal.
  - For example, consider the discrete uniform probability distribution defined as X={1,2,3,4,5}, p(x)=0.2.
    - ✓Here, the probability for each outcome is equal to 0.2 [i.e., p(1)=p(2)=p(3)=p(4)=p(5)=0.2].

R9 Common Probability Distributions

- **Binomial distribution**
  - Bernoulli random variable

    P(Y=1)=p　P(Y=0)=1-p

  - Binomial random variable，the probability of x successes in n trails

$$p(x) = P(X = x) = \binom{n}{x} p^{x}(1-p)^{n-x}$$

  - Expectations and variances

	Expectation	Variance
Bernoulli random variable (Y)	p	p(1-p)
Binomial random variable (X)	np	np(1-p)

R9 Common Probability Distributions

- **Continuous Uniform Distribution**

---- is defined over a range that spans between some lower limit, a, and upper limit, b, which serve as the parameters of the distribution.

- Properties of Continuous uniform distribution
  - For all $a \le x1 < x2 \le b$
  - $P(X<a \text{ or } X>b) = 0$
  - $P(x_1 \le X \le x_2) = (x_2 - x_1)/(b - a)$

R9 Example: Common Probability Distributions

1. Which of the following statements about probability distributions is **TRUE**?
   A. A continuous uniform distribution has a lower limit but no upper limit.
   B. A cumulative distribution function defines the probability that a random variable is greater than a given value.
   C. A binomial distribution counts the number of successes that occur in a fixed number of independent trials that have mutually exclusive (i.e. yes or no) outcomes.

- Correct answer: C

2. A random variable with a finite number of equally likely outcomes is best described by a:
   A. Binomial distribution.
   B. Bernoulli distribution.
   C. Discrete uniform distribution.

- Correct answer: C

R9 Example: Common Probability Distributions

3. A recent study indicated that 60% of all businesses have a fax machine. From the binominal probability distribution table, the probability that exactly four businesses will have a fax machine in a random selection of six businesses is:

   A. 0.138

   B. 0.276

   C. 0.311

- Correct answer: C

4. Assume that 40% of candidates who sit for the CFA examination pass it the first time. Of a random sample of 15 candidates who are sitting for the exam for the first time, what is the expected number of candidates that will pass?

   A. 0.375

   B. 4.000

   C. 6.000

- Correct answer: C

R9 Example: Common Probability Distributions

5. An analyst has recently determined that only 60 percent of all U.S. pension funds have holdings in hedge funds. In evaluating this probability, a random sample of 50 U.S. pension funds is taken. The number of U.S. pension funds in the sample of 50 that have hedge funds in their portfolio would most accurately be described as:
   A. A binomial random variable.
   B. A Bernoulli random variable.
   C. A continuous random variable.

➢ Correct answer: A

6. An energy analyst forecasts that the price per barrel of crude oil five years from now will range between USD$75 and USD$105. Assuming a continuous uniform distribution, the probability that the price will be less than USD$80 five years from now is closest to:
   A. 5.6%.
   B. 16.7%.
   C. 44.4%.

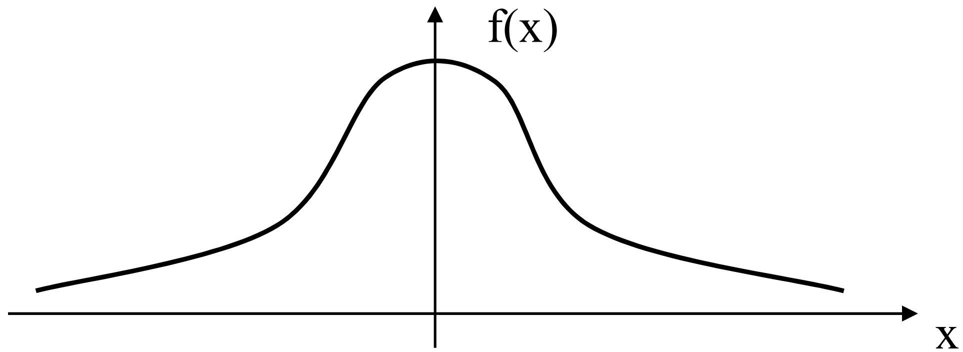
➢ Correct answer: B

R9 Common Probability Distributions

- Tracking error is the difference between the total return on a portfolio and the total return on the benchmark against which its performance is measured.

R9 Common Probability Distributions

➤ **The shape of the density function**

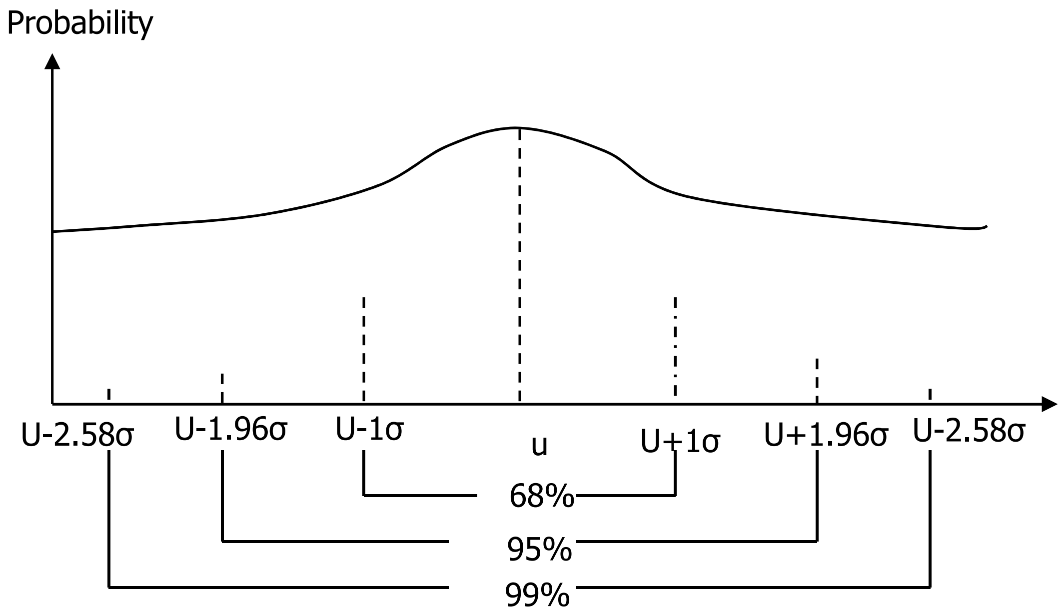


➤ **Properties**:

- X~N(μ , σ²)
- Symmetrical distribution: skewness=0; kurtosis=3
- A linear combination of normally distributed random variables is also normally distributed.
- The tails get thin and go to zero but extend infinitely, asympotic（渐近）

R9 Common Probability Distributions

- **The confidence intervals**
  - 68% confidence interval is $[\mu-\sigma,\mu+\sigma]$
  - 90% confidence interval is $[\mu-1.65\sigma,\mu+1.65\sigma]$
  - 95% confidence interval is $[\mu-1.96\sigma,\mu+1.96\sigma]$
  - 99% confidence interval is $[\mu-2.58\sigma,\mu+2.58\sigma]$



R9 Example: Common Probability Distributions

- The average return of a mutual fund is 10.5% per year and the standard deviation of annual returns is 18%. If returns are approximately normal, what is the 95% confidence interval for the mutual fund return next year?
- **Answer:**

Here μ and σ are 10.5% and 18%, respectively. Thus, the 95% confidence interval for the return, R, is:

$10.5 \pm 1.96(18) = -24.78\%$ to 45.78%

Symbolically, this result can be expressed as:

$P(-24.78 < R < 45.78) = 0.95$ or 95%

The interpretation is that the annual return is expected to be within this interval 95% of the time, or 95 out of 100 years.

R9 Example: Common Probability Distributions

- An analyst determined that approximately 99 percent of the observations of daily sales for a company were within the interval from $230,000 to $480,000 and that daily sales for the company were normally distributed. The mean daily sales and standard deviation of daily sales, respectively, for the company were closest to:

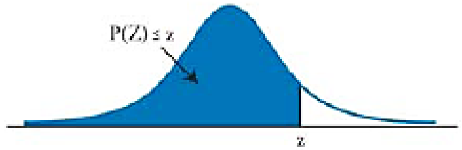
	Mean daily sales	Standard deviation of daily sales
A.	$351,450	$48,450
B.	$351,450	$83,333
C.	$355,000	$48,450

- Correct answer: C

R9 Common Probability Distributions

- **Standard normal distribution**
  - N(0,1) or Z
  - Standardization: if X~N(μ , σ²), then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
  - Z-table
- F(-z)=1-F(z)
- P(Z>z) = 1 –F(z)

Cumulative Z-Table



Standard Normal Distribution

$P(Z \leq z) = N(z)$ for $z \geq 0$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

R9 Example: Common Probability Distributions

1. Based on a normal distribution with a mean of 500 and a standard deviation of 150, the z-value for an observation of 200 is closest to:

   A. –2.00.

   B. –1.75.

   C. 1.75.

- Correct answer: A

2. For a standard normal distribution, F(0) is:

   A. 0.0

   B. 0.1

   C. 0.5

- Correct answer: C

R9 Example: Common Probability Distributions

- A study of hedge fund investors found that their annual household incomes are normally distributed with a mean of $175.000 and a standard deviation of $25,000. F(1)=0.8413, F(2)=0.9772, F(3)=0.9987

1. The percent of hedge fund investors that have incomes less than $100,000 is closest to:
   A. 0.05%
   B. 0.10%
   C. 0.13%
2. The percent of hedge fund investors that have incomes greater than $225,000 is closest to:
   A. 0.50%
   B. 1.10%
   C. 2.28%
3. The percent of hedge fund investors that have incomes greater than $150,000 is closest to:
   A. 34.13%
   B. 68.26%
   C. 84.13%

R9 Common Probability Distributions

- **Shortfall risk:** R_L= threshold level return, minimum return required
  - Minimize ($Rp< R_L$)
- **Roy's safety-first criterion**

$$[E(R_P) - R_L]/\sigma_P$$

- **Maximize S-F-Ratio**
  - Maximize $SFR=\dfrac{E(R_P)-R_L}{\sigma_P}$ <=> Minimize P ($Rp< R_L$)

R9 Example: Common Probability Distributions

- A portfolio manager gathered the following information about four possible asset allocations:

Allocation	Expected annual return	Standard deviation of return
A	10%	6%
B	25%	14%
C	18%	17%

- The manager's client has stated that her minimum acceptable return is 8%. Based on Roy's safety-first criterion, the *most* appropriate allocation is:
  - A. Allocation A.
  - B. Allocation B.
  - C. Allocation C.
- Correct answer: B

R9 Example: Common Probability Distributions

- You are researching asset allocations for a client with an \$800,000 portfolio. Although her investment objective is long-term growth, at the end of a year she may want to liquidate \$30,000 of the portfolio to fund educational expenses. If that need arises, she would like to be able to take out he \$30,000 without invading the initial capital of \$80,000. The following table shows three alternative allocations.

	A	B	C
Expected annual return	25	11	14
Standard deviation of return	27	8	20

- Address these questions (assume normality for Parts 2 and 3):

1. Given the client's desire not to invade the \$800,000 principal, what is the shortfall level, R_L? Use this shortfall level to answer Part 2.
2. According to the safety-first criterion, which of the tree allocations is the best?
3. What is the probability that the return on the safety-first optimal portfolio will be less than the shortfall level? (F(0.91)=0.8186)

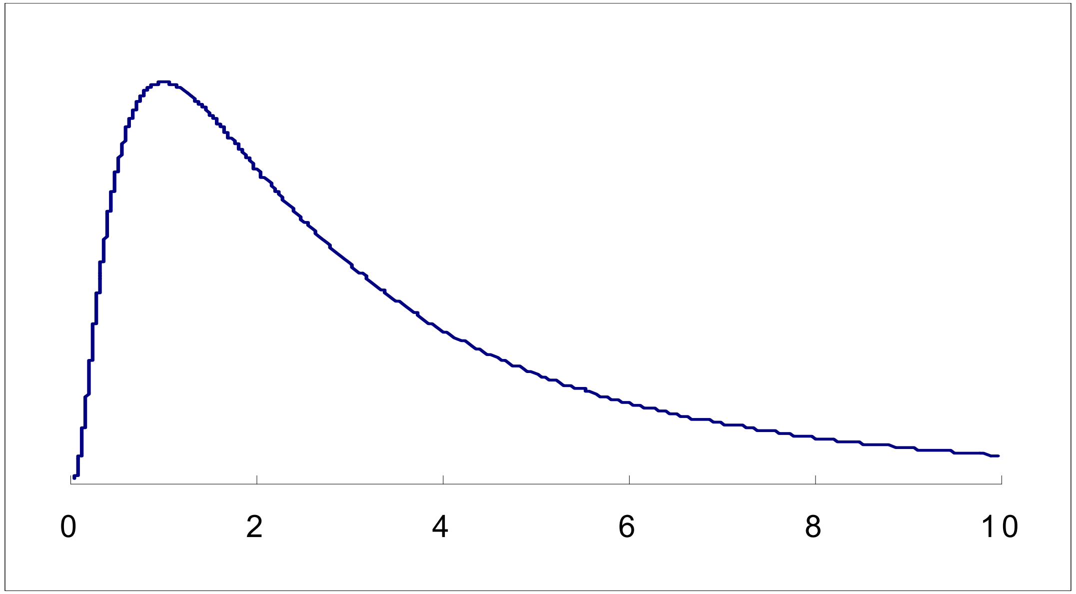
- **Solution:**

1. R_L=30,000/800,000=3.75%
2. A: SFR_A=(25-3.75)/27=0.79; B: SFR_B=(11-3.75)/8=0.91; C: SFR_C=(14-3.75)/20=0.5125; B is best.
3. $P(R_B<3.75)=P[(R_B-11)/8 < (3.75-11)/8] = F(-0.91) = 1-F(0.91)=1-0.8186=0.1814$

   The safety-first optimal portfolio has a roughly 18% chance of not meeting a 3.75% return threshold.

R9 Common Probability Distributions

- **Definition:** If lnX is normal, then X is lognormal, which is used to describe the price of asset
- **Right skewed**
- **Bounded from below by zero**



R9 Example: Common Probability Distributions

1. Compared to a normal distribution, a lognormal distribution is ***least likely*** to be:
   A. Skewed to the left.
   B. Skewed to the right.
   C. Useful in describing the distribution of stock prices.

➢ Correct answer: A

2. An analyst stated that lognormal distribution are suitable for describing asset returns and that normal distributions are suitable for describing distributions of asset prices. Is the analyst's statement correct with respect to:

	Lognormal distribution	Normal distribution
A.	No	No
B.	No	Yes
C.	Yes	No

➢ Correct answer: A

R9 Common Probability Distributions

- Discrete: $$EAY = (1+\frac{R}{m})^m - 1$$
- Continuous: $$EAR = \lim_{m\to\infty}(1+\frac{R}{m})^m - 1 = e^R - 1$$
- $$\frac{S_1}{S_0} = 1+HPR = e^{R_{CC}} \text{（持有一年）}$$
- $$1+HPR_T = e^{R_{CC}\times T} \text{（持有T年）}$$

R9 Common Probability Distributions

- Monte Carlo simulation vs Historical simulation
  - Monte Carlo simulation uses randomly generated values for risk factors, based on their assumed distributions, to produce a distribution of possible security values, to analyze the complex instrument;
    - ✓ Limitations:
      - ◆ It is fairly complex and will assume a parameter distribution.
      - ◆ It is not an analytic method but a statistical one, and cannot provide the insights that analytic methods can.
  - Historical simulation uses randomly selected past changes in these risk factors to generate a distribution of possible security values, can't answer the "What-If".
    - ✓ Limitations: the past can not indicate the future and historical simulation cannot address the sort of "what if" questions that Monte Carlo simulation can.

R9 Example: Common Probability Distributions

- Monte Carlo simulation is best described as:

  A. An approach to back testing data

  B. A restrictive form of scenario analysis

  C. Providing a distribution of possible solutions to complex functions

- Solution: C

R10 Sampling and Estimation

- Sampling and Estimation
  - Simple random and stratified random sampling, time-series and cross-sectional data
  - Central limit theorem
  - Standard error of the sample mean的意义及计算
  - The desirable properties of an estimator
  - Student’s t-distribution的特点
  - Criteria for selecting the appropriate test statistic，计算confidence interval
  - Five kinds of biases

R10 Sampling and Estimation

- Sampling and estimation
  - Simple random sampling
  - Stratified random sampling: to separate the population into smaller groups based on one or more distinguishing characteristics. Stratum and cells=M*N
- Sampling error: sampling error of the mean= sample mean- population mean
- The sample statistic itself is a random variable and has a probability distribution.

R10 Sampling and Estimation

- Time-series data
  - consist of observations taken over a period of time at specific and equally spaced time intervals.
- Cross-sectional data
  - a sample of observations taken at a single point in time.

Time-series data	Cross-sectional data
a collection of data recorded over a period of time	a collection of data taken at a single point of time.

R10 Example: Sampling and Estimation

- Greg Goldman, research analyst in the fixed-income area of an investment bank, needs to determine the average duration of a sample of twenty 15-year fixed-coupon investment grade bonds. Goldman first categorizes the bonds by risk class and then randomly selects bonds from each class. After combining the bonds selected (bond ratings and other information taken as of March 31st of the current year), he calculates a sample mean duration of 10.5 years.
- Assuming that the actual population mean is 9.7 years, which of the following statements about Goldman's sampling process and sample is **FALSE**?
  A. Goldman used stratified random sampling.
  B. The sampling error of the means equals 0.8 years.
  C. Goldman is using time-series data.
- Correct answer: C

R10 Sampling and Estimation

Central Limit Theory

- For simple random samples of size n from a population with a mean μ and a variance σ^2 but without known distribution, the <u>sampling distribution of the sample mean</u> approaches <u>N(μ , σ^2/n)</u> if the sample size is sufficiently large ($n \geq 30$).

条件：1. $n \geq 30$　　2. 总体均值、方差已知

结论: 1.服从正态分布　2. $\mu_{population} = \mu_{sample}$　$s^2 = \sigma^2 / n$

R10 Example: Sampling and Estimation

- If the distribution of the population from which the samples are drawn is positively skewed, and given that the sample size is large, the sampling distribution of the sample means is most likely:

  A. approximately normally distributed.

  B. to have a variance equal to that of the entire population.

  C. to have a mean smaller than the mean of the entire population.

- Solution: A

R10 Sampling and Estimation

Standard error of the sample mean

- Known population variance $\sigma_{\bar{x}} = \sigma / \sqrt{n}$
- Unknown population variance $s_{\bar{x}} = s / \sqrt{n}$

R10 Example: Sampling and Estimation

- An analyst gathered the following information:

Sample mean	12%
Sample size	50
Sample variance	$30(\%)^2$

- The standard error of the sample mean is *closest* to:

  A. 0.47%.

  B. 0.64%.

  C. 0.77%.

- Correct answer: C

R10 Sampling and Estimation

- The desirable properties of an estimator:
  - **Unbiasedness:** expected value of the estimator is equal to the parameter that are trying to estimate
  - **Efficiency:** for all unbiased estimators, if the sampling dispersion is smaller than any other unbiased estimators, then this unbiased estimator is called efficient.
  - **Consistency:** the accuracy of the parameter estimate increases as the sample size increases. **(the standard deviation of the parameter estimate decreases as the sample size increases)**
    - ✓ As the sample size increases, the standard error of the sample mean falls.

R10 Example: Sampling and Estimation

- Shawn Choate is thinking about his graduate thesis. Still in the preliminary stage, he wants to choose a variable of study that has the most desirable statistical properties. The statistic he is presently considering has the following characteristics:
  - The expected value of the sample mean is equal to the population mean.
  - The variance of the sampling distribution is smaller than that for other estimators of the parameter.
  - As the sample size increases, the standard error of the sample mean rises and the sampling distribution is centered more closely on the mean.
- Select the *best* choice. Choate's estimator is:
  A. Unbiased, efficient, and consistent.
  B. Efficient and consistent.
  C. Unbiased and efficient.
- Correct answer: C

R10 Sampling and Estimation

- **Point estimate:** the statistic, computed from sample information, which is used to estimate the population parameter
- **Confidence interval estimate:** a range of values constructed from sample data so the parameter occurs within that range at a specified probability. α—the level of significance
- **Interval Estimation（also see Chapter: Hypothesis Testing ）**
  - Level of significance （alpha）
  - Degree of Confidence （1－alpha）
  - Confidence Interval = [Point Estimate +/- (reliability factor) * Standard error]

R10 Example: Sampling and Estimation

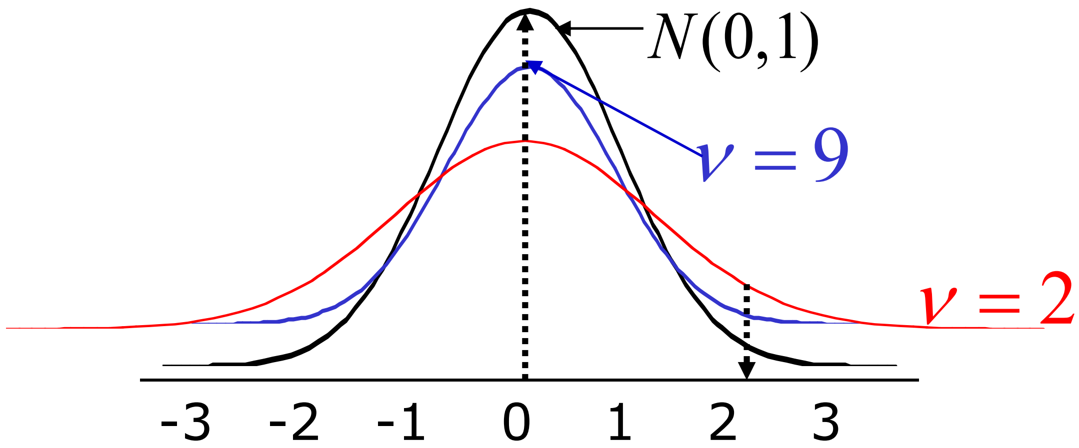
- The width of a confidence interval *most likely* will be smaller if the sample variance and number of observations, respectively, are:

	Sample variance	Number of observations
A.	Smaller	Smaller
B.	Smaller	Larger
C.	Larger	Smaller

- Correct answer: B

R10 Sampling and Estimation

- **Student's t-distribution**
  - Symmetrical
  - Degrees of freedom (df): n-1
  - Less peaked than a normal distribution ("fatter tails")
  - As the degrees of freedom gets larger, the shape of t-distribution approaches standard normal distribution



R10 Example: Sampling and Estimation

- An analyst stated that as degrees of freedom increase, a t-distribution will become more peaked and the tails of the t-distribution will become less fat. Is the analyst's statement correct with respect to the t-distribution:

	Become more peaked?	Tails becoming less fat?
A.	No	No
B.	No	Yes
C.	Yes	Yes

- Correct answer: C

R10 Sampling and Estimation

$$\boxed{\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

$$\boxed{\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}}$$

When sampling form a:	Test Statistic	
	small sample (n<30)	large sample (n>=30)
Normal distribution with known variance	z- Statistic	z- Statistic
Normal distribution with unknown variance	t- Statistic	t- Statistic/z
Nonnormal distribution with known variance	not available	z- Statistic
Nonnormal distribution with unknown variance	not available	t- Statistic/z

R10 Example: Sampling and Estimation

1. What is the most appropriate test statistic for constructing confidence intervals for the population mean when the population mean when the population is normally distributed, but the variance is unknown?
   A. The z-statistic at α with n degrees of freedom
   B. The t-statistic at α/2 with n degrees of freedom
   C. The t-statistic at α/2 with n-1 degrees of freedom

➢ **Answer: C**

2. When constructing a confidence interval for the population mean of nonnormal distribution when the population variance is unknown and the sample size large (n>30), an analyst may acceptably use:
   A. Either a z-statistic or a t-statistic
   B. Only a z-statistic or α with n degrees of freedom
   C. Only a t-statistic or α/2 with n degrees of freedom

➢ **Answer: A**

R10 Example: Sampling and Estimation

3. The 95% confidence interval of the sample mean of employee age for a major corporation is 19 years to 44 years based on a z-statistic. The population of employees is more than 5,000 and the sample size of this test is 100. Assuming the population is normally distribution, the standard error of mean employee age is closest to:

   A. 1.96

   B. 2.58

   C. 6.38

- **Answer: C**

R10 Sampling and Estimation

- **Data-mining bias**
  - Refers to results where the statistical significance of the pattern is overestimated because the results were found through data mining.
- **Sample selection bias**
  - Some data is systematically excluded from the analysis, usually because of the lack of availability.
- **Survivorship bias**
  - Usually derives from sample selection for only the existing portfolio are included
- **Look-ahead bias**
  - Occurs when a study tests a relationship using sample data that was not a available on the test date.
- **Time-period bias**
  - Time period over which the data is gathered is either too short or too long. If the time period is too short, research results may reflect phenomena specific to that time period, or perhaps even data mining.

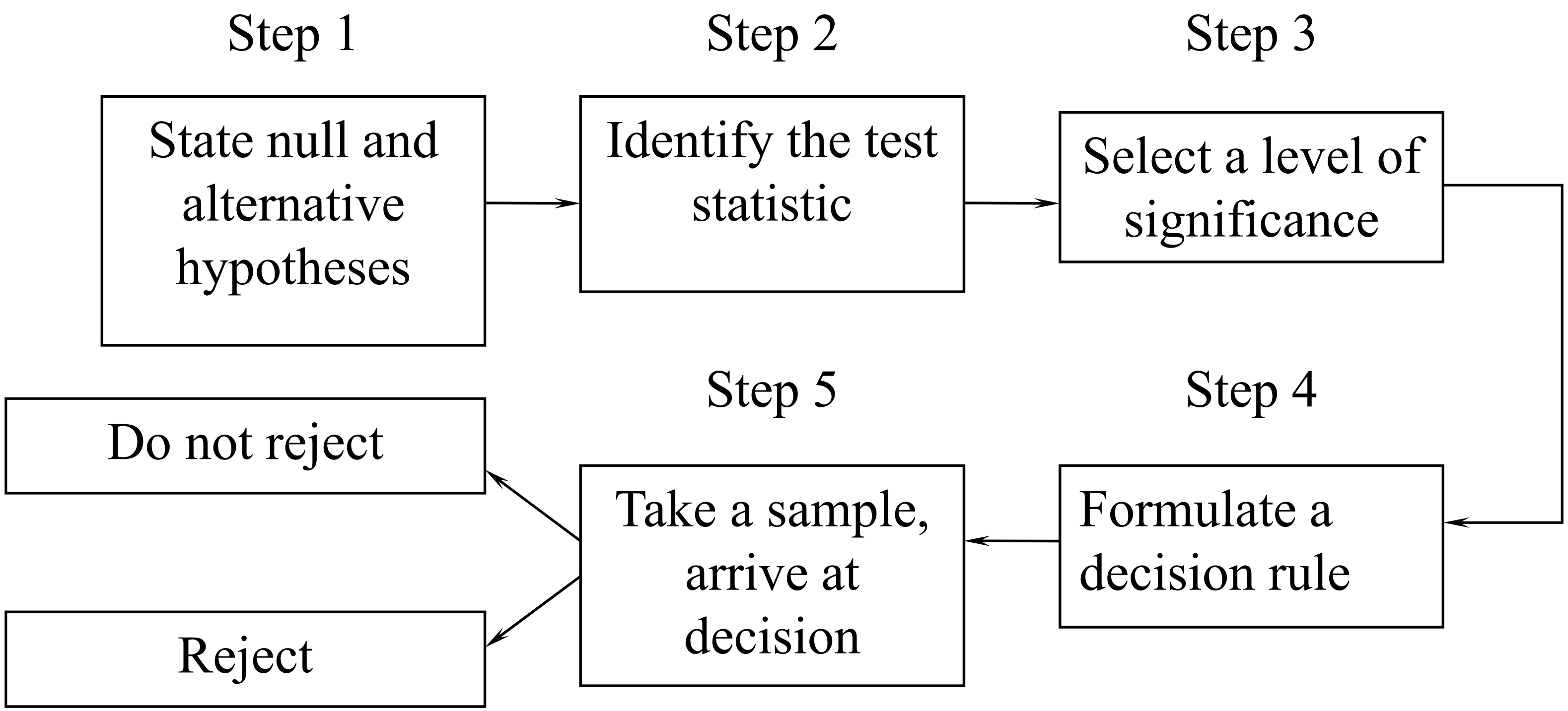
R10 Example: Sampling and Estimation

- Sunil Hameed is a reporter with the weekly periodical The Fun Finance Times. Today, he is scheduled to interview a researcher who claims to have developed a successful technical trading strategy based on trading on the CEO's birthday (sample was taken from the Fortune 500). After the interview, Hameed summarizes his notes (partial transcript as follows). The researcher:
  - Used the same database of data for all his tests and has not tested the trading rule on out-of-sample data.
  - Excluded stocks for which he could not determine the CEO's birthday.
- Select the choice that *best* completes the following: Hameed concludes that the research is flawed because the data and process are biased by:
  A. Data mining and sample selection bias.
  B. Data mining and look-ahead bias.
  C. Time-period bias and survivorship bias.
- Correct answer: A

R11 Hypothesis Testing

- Hypothesis testing
  - The steps of hypothesis testing
  - The null hypothesis and alternative hypothesis, one-tailed and two-tailed test
  - Test statistics的选择和计算
  - Type I and type II errors
  - Decision rule
  - The Chi-square test and F-test
  - Parameter tests and non-parameter tests

R11 Hypothesis Testing



R11 Hypothesis Testing

- **Define Hypothesis**

Statistical assessment of a statement or idea regarding a population parameter.

Null hypothesis and Alternative hypothesis (we want to assess)

$H_0 : \mu = \mu_0 \qquad H_a : \mu \neq \mu_0$

- The fact we suspect and want to reject
- For population not sample

R11 Hypothesis Testing

- **One-tailed and Two-tailed tests of Hypothesis**

Two-tailed	$H_0: \mu = \mu_0$	$H_a: \mu \neq \mu_0$
One-tailed	$H_0: \mu \leq \mu_0$	$H_a: \mu > \mu_0$
	$or, H_0: \mu \geq \mu_0$	$H_a: \mu < \mu_0$

R11 Example: Hypothesis Testing

- In the hypothesis testing, assess whether if mean excess the benchmark, how to set the null hypothesis?

  A. $\mu \prec \mu_0$

  B. $\mu \leq \mu_0$

  C. $\mu \succ \mu_0$

- Correct answer: B

R11 Example: Hypothesis Testing

1. Austin Roberts believes that the mean price of houses in the area is greater the \$145,000. the appropriate alternative hypothesis is:
   A. H_a: μ < \$145,000
   B. H_a: $\mu \geq$ \$145,000
   C. H_a: μ > \$145,000

- **Answer: C**

2. An analyst is conducting a hypothesis test to determine if the mean time spent on investment research is different from three hours per day. The appropriate null hypothesis for the described test is:
   A. H_0: μ = 3 hours, two-tailed test.
   B. H_0: μ = 3 hours, one-tailed test.
   C. H_0: $\mu \geq$ 3 hours, two-tailed test.

- **Answer: A**

R11 Hypothesis Testing

➢ **Test statistic**

$$Test\ Statistic = \frac{Sample\ statistics - Hypothesized\ value}{stanard\ error\ of\ the\ sample\ statistic}$$

- Test Statistic follows Normal, T, Chi Square or F distributions
- Test Statistic has formula. Calculate it with the sample data. We should emphasize Test Statistic is calculated by ourselves not from the table.
- This is the general formula but only for Z and T distribution.

Examples:

$$Test\ Statistic = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \qquad Test\ Statistic = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$

R11 Hypothesis Testing

- **Critical value (关键值，实际就是分位数）**
  - Found in the Z, T, Chi Square or F distribution tables not calculated by us
  - Under given one tailed or two tailed assumption, critical value is determined solely by the significance level.
- **Decision rule**
  - **Critical Value Method**

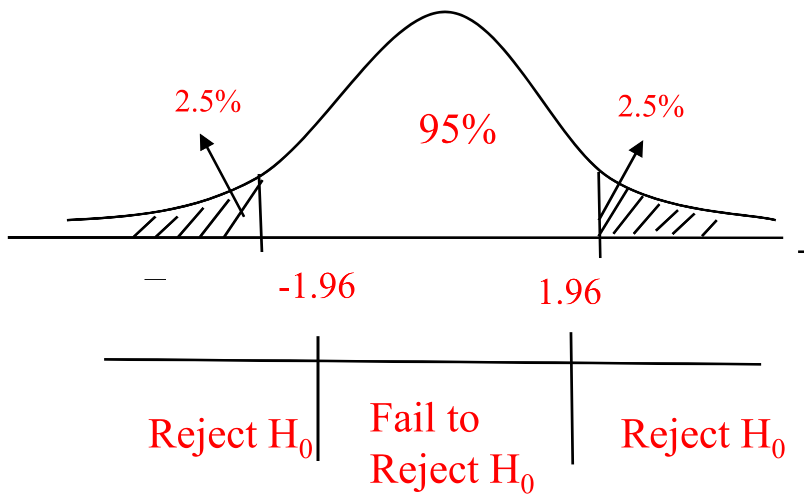
Significance Level?

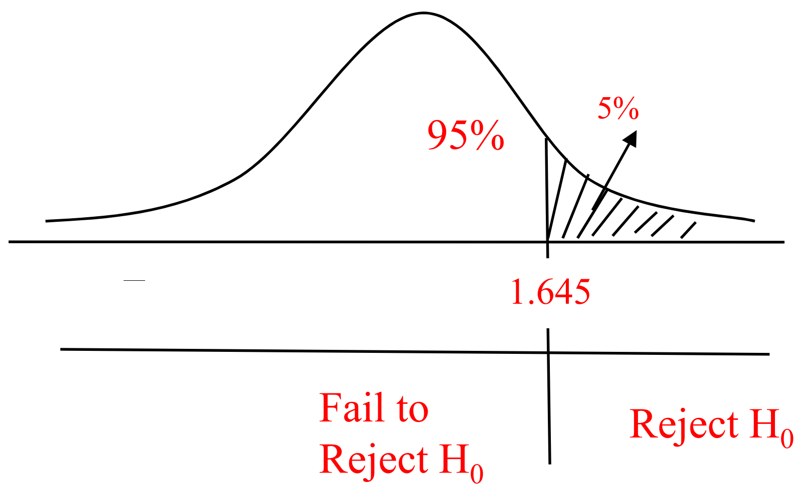
Two tailed or one tailed test?

Reject region? Critical Value under the condition

Compare the Test Statistic and Critical Value

R11 Hypothesis Testing





- Reject H_0 if |test statistic|>critical value
- Fail to reject H_0 if |test statistic|<critical value
- **Statement**

  - cannot say "accept the null hypothesis", only can say "cannot reject"
  - ***** is significantly different from ******
  - *****is not significantly different from ******

R11 Hypothesis Testing

- **Relation between Confidence Intervals and Hypothesis Tests**

- Confidence Interval = [sample statistic ± (critical value)(standard error)]
- Center of Interval = sample statistic
- Length of Interval = 2*(critical value)(standard error)

R11 Example: Hypothesis Testing

- **Example: Two-tailed test**
- A researcher has gathered data on the daily returns on a portfolio of call options over a recent 250-day period. The mean daily return has been 0.1%, and the sample standard deviation of daily portfolio returns is 0.25%. The researcher believes that the mean daily portfolio return is not equal to zero. Construct a hypothesis test of the researcher's belief.
- **Answer:**
  - First we need to specify the null and alternative hypotheses. The null hypothesis is the one the researcher expects to reject.

$$H_0 : \mu_0 = 0 \text{ versus } H_a : \mu_0 \neq 0$$

  - Since the null hypothesis is an equality, this is a two-tailed test. At a 5% level of significance, the critical z-values for a two-railed test are ± 1.96, so the decision rule can be stated as:
  - Reject H_0 if test statistic < -1.96 or test statistic > +1.96

R11 Example: Hypothesis Testing

- The standard error of the sample mean is the adjusted standard deviation of the sample. When the sample statistic is the sample mean, x, the standard error of the sample statistic for sample size n is calculated as:

$$S_{\overline{X}} = \frac{S}{\sqrt{n}}$$

- Since our sample statistic here is a sample mean, the standard error of the sample mean for a sample size of 250 is $0.0025/\sqrt{250}$ and our test statistic is:

$$\frac{0.001}{\left(\frac{0.0025}{\sqrt{250}}\right)} = \frac{0.001}{0.000158} = 6.33$$

- Since 6.33 > 1.96, we reject the null hypothesis that the mean daily option return is equal to zero. Note that when we reject the null, we conclude that the sample value is significantly different from the hypothesized value. We are saying that the two values are different from one another after considering the variation in the sample. That is, the mean daily return of 0.001 is statistically different from zero given the sample's standard deviation and size.

R11 Example: Hypothesis Testing

- **Example: One-tailed test**
- Perform a z-test using the option portfolio data from the previous example to test the belief that option returns are positive.
- **Answer:**
  - In this case, we use a one-tailed test with the following structure:

    Ho: $\mu o \leq 0$ versus Ha: $\mu > 0$
  - The appropriate decision rule for this one-tailed z-test at a significance level of 5% is:

    Reject Ho if test statistic > 1.65
  - The test statistic is computed the same way, regardless of whether we are using a one-tailed or two-tailed test. From the previous example, we know that the test statistic for the option return sample is 6.33. Since 6.33 > 1.645, we reject the null hypothesis and conclude that mean returns are statistically greater than zero at a 5% level of significance.

R11 Example: Hypothesis Testing

- An analyst conducts a two-tailed test to determine if earnings estimates are significantly different from reported earnings. The sample size was over 100. The computed Z-statistic is 1.25. Using a 5 percent confidence level, which of the following statements is **TRUE**?
  - A. Both the null and the alternative are significant.
  - B. You cannot determine what to do with the information given.
  - C. Fail to reject the null hypothesis and conclude that the earnings estimates are not significantly different from reported earnings.
- **Correct answer: C**

R11 Hypothesis Testing

- **P-value Method**
  - The **p-value** is the smallest level of significance at which the null hypothesis can be reject
  - p-value < α: reject H_0; p-value > α: do not reject H_0.
  - P↓, easier to reject H_0
- **Example:**
- The p-value for a two-tailed test of sample mean is 1.68%. Which of the following is true?
  A. We can reject the null with 95% confidence
  B. We can reject the null with 99% confidence
  C. the largest probability of rejecting the null hypothesis is 1.68%
- **Answer: A**

R11 Hypothesis Testing

- **Type I error and Type II error**
  - **Type I error**: 拒真，reject the null hypothesis when it's actually true
    - ✓ Significance level (α): the probability of making a Type I error
    - ✓ Significance level =P(Type I error)=P(H_0×|H_0√)
  - **Type II error**: 取伪，fail to reject the null hypothesis when it's actually false
    - ✓ P(Type II error)=P(H_1×|H_1√)
    - ✓ Power of a test: the probability of correctly rejecting the null hypothesis when it is false
    - ✓ Power of a test = 1−P(Type II error)= P(H_1√|H_1√)

R11 Hypothesis Testing

Decision	True condition	
	H_0 **is true**	H_0 **is false**
Do not reject H_0	**Correct Decision**	**Incorrect Decision** Type II error
Reject H_0	**Incorrect Decision** Significance level α =P (Type I error)	**Correct Decision** Power of test = 1- P (Type II error)

- With other conditions unchanged, either error probability arises at the cost of the other error probability decreasing.
- How to reduce both errors? Increase the Sample Size.

R11 Example: Hypothesis Testing

1. Kyra Mosby, M.D., has a patient who is complaining of severe abdominal pain. Based on an examination and the results from laboratory tests, Mosby states the following diagnosis hypothesis: Ho: Appendicitis, HA: Not Appendicitis. Dr. Mosby removes the patient's appendix and the patient still complains of pain. Subsequent tests show that the gall bladder was causing the problem. By taking out the patient's appendix, Dr. Mosby:
   A. Made a Type I error.
   B. Is correct.
   C. Made a Type II error.

➢ Correct answer: C

2. If the sample size increases, the probability of get the Type Ⅰ and Type Ⅱ error will

	Type Ⅰ	Type Ⅱ
A.	increase	increase
B.	not change	not change
C.	decrease	decrease

➢ Correct answer: C

R11 Example: Hypothesis Testing

3. All else equal, is specifying a larger significance level in a hypothesis test likely to increase the probability of a:

	Type I error?	Type II error?
A.	No	No
B.	No	Yes
C.	Yes	No

- Correct answer: C

4. What is the definition of the power test? Power test is the probability to:
   A. Reject the true null hypothesis while it is true
   B. Reject the false null hypothesis while it is indeed false
   C. Can not reject the true hypothesis

- Correct answer: B

R11 Hypothesis Testing

➤ **Test Population Mean**

1. One normal population with **known variance** $\boxed{\textbf{Z distribution}}$

2. One normal population with **unknown variance**

	Normal population, n<30	**n>30**
Known variance	**z-test**	**z-test**
Unknown variance	**t-test**	**t-test or z-test**

R11 Summary of Hypothesis Testing

Test type	Assumptions	H_0	Test-statistic	Critical value
Mean hypothesis testing	Normally distributed population, known population variance	$\mu=0$	$Z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	N(0,1)
	Normally distributed population, unknown population variance	$\mu=0$	$t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	t(n-1)
	Independent populations, unknown population variances assumed equal	$\mu_1-\mu_2=0$	t	$t(n_1+n_2-2)$
	Independent populations, unknown population variances not assumed equal	$\mu_1-\mu_2=0$	t	t
	Samples not independent, paired comparisons test	$\mu_d=0$	$t=\bar{d}/s_{\bar{d}}$	t(n-1)
Variance hypothesis testing	Normally distributed population	$\sigma^2=\sigma_0^2$	$\chi^2=\frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2(n-1)$
	Two independent normally distributed populations	$\sigma_1^2=\sigma_2^2$	$F=s_1^2/s_2^2$	$F(n_1-1,n_2-1)$

R11 Example: Hypothesis Testing

➢ **Paired comparisons test**

An analyst collects the following data related to paired observations for Sample A and Sample B. Assume that both samples are drawn from normally distributed populations and that the population variances are not known.

Paired Observation	Sample A Value	Sample B Value
1	25	18
2	12	9
3	–5	–8
4	6	3
5	–8	1

The *t*-statistic to test the hypothesis that the mean difference is equal to zero is *closest* to:

A. 0.23.
B. 0.27.
C. 0.52.

Answer = C

R11 Example: Hypothesis Testing

- **Example: Chi-square test for a single population variance**
- Historically, High-Return Equity Fund has advertised that its monthly returns have a standard deviation equal to 4%. This was based on estimates from the 1990-1998 period. High-Return wants to verify whether this claim still adequately describes the standard deviation of the fund's returns. High-Return collected monthly returns for the 24-month period between 1998 and 2000 and measured a standard deviation of monthly returns of 3.8%. Determine if the more recent standard deviation is different from the advertised standard deviation.
- Answer:
  - *State the hypothesis.* The null hypothesis is that the standard deviation is equal to 4% and, therefore, the variance of monthly returns for the population is $(0.04)^2$=0.0016. Since High-Return simply wants to test whether the standard deviation has changed, up or down, a two-sided test should be used The hypothesis test structure takes the form:
  $$H_0 : \sigma_0^2 = 0.0016 \text{ versus } H_a : \sigma_0^2 \neq 0.0016$$
  - *Select the appropriate test statistic.* The appropriate test statistic for tests of variance using the chi-square distribution is computed as follows:
  $$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

R11 Example: Hypothesis Testing

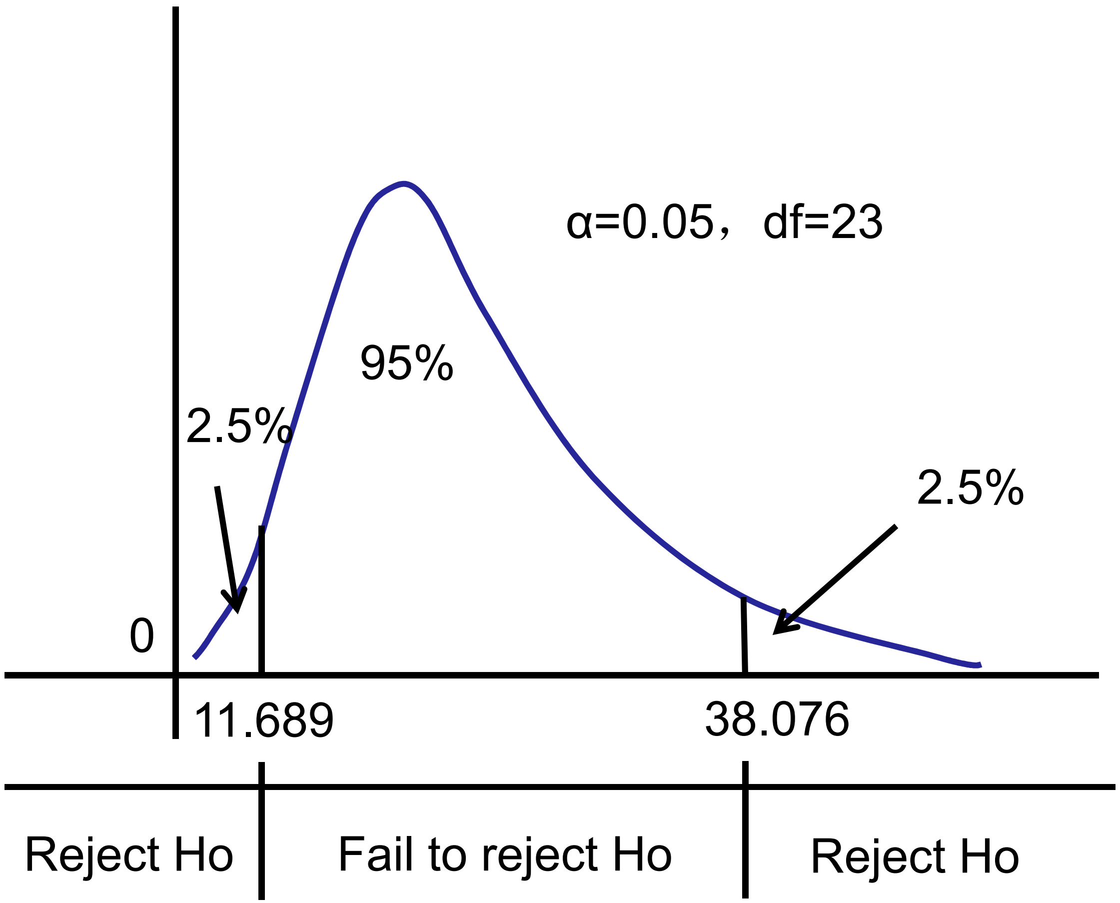
- *Specify the level of significance.* Let's use a 5% level of significance, meaning there will be 2.5% probability in each tail of the chi-square distribution.
- *State the decision rule regarding the hypothesis.* With a 24-month sample, there are 23 degrees of freedom. Using the table of chi-square values at the back of this book, for 23 degrees of freedom and probabilities of 0.975 and 0.025, we find two critical values, 11.689 and 38.076. Thus, the decision rule is:
- Reject H_0 if

$$\chi^2 < 11.689, \text{ or } \chi^2 > 38.076$$

- This decision rule is illustrated in the following figure.

R11 Example: Hypothesis Testing

Decision Rule for a Two-Tailed Chi-Square Test of a Single Population Variance



R11 Example: Hypothesis Testing

- *Collect the sample and calculate the sample statistics*. Using the information provided, the test statistic is computed as:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(23)(0.001444)}{0.0016} = \frac{0.033212}{0.0016} = 20.7575$$

- *Make a decision regarding the hypothesis*. Since the computed test statistic, χ^2, falls between the two critical values, we fail to reject the null hypothesis that the variance is equal to 0.0016.

- *Make a decision based on the results of the test*. It can be concluded that the recently measured standard deviation is close enough to the advertised standard deviation that we cannot say that it is different from 4%, at a 5% level of significance.

R11 Example: Hypothesis Testing

- **Example: F-test for equal variances**
- Annie Cower is examining the earnings for two different industries. Cower suspects that the earnings of the textile industry are more divergent than those of the paper industry. To confirm this suspicion, Cower has looked at a sample of 31 textile manufacturers and a sample of 41 paper companies. She measured the sample standard deviation of earnings across the textile industry to be $4.30 and that of the paper industry companies to be $3.80. Determine if the earnings of the textile industry have greater standard deviation than those of the paper industry.
- **Answer:**
  - *State the hypothesis*. In this example, we are concerned with whether the variance of the earnings of the textile industry is greater (more divergent) than the variance of the earnings of the paper industry. As such, the test hypotheses can be appropriately structured as:

$$\mathrm{H}_0 : \sigma_1^2 \le \sigma_2^2 \text{ versus } \mathrm{H}_a : \sigma_1^2 > \sigma_2^2$$

  - where:
  - σ_1^2 = variance of earnings for the textile industry
  - σ_2^2 = variance of earnings for the paper industry
  - Note: $\sigma_1^2 > \sigma_2^2$

R11 Example: Hypothesis Testing

- *Select the appropriate test statistic*. For tests of difference between variances, the appropriate test statistic is:

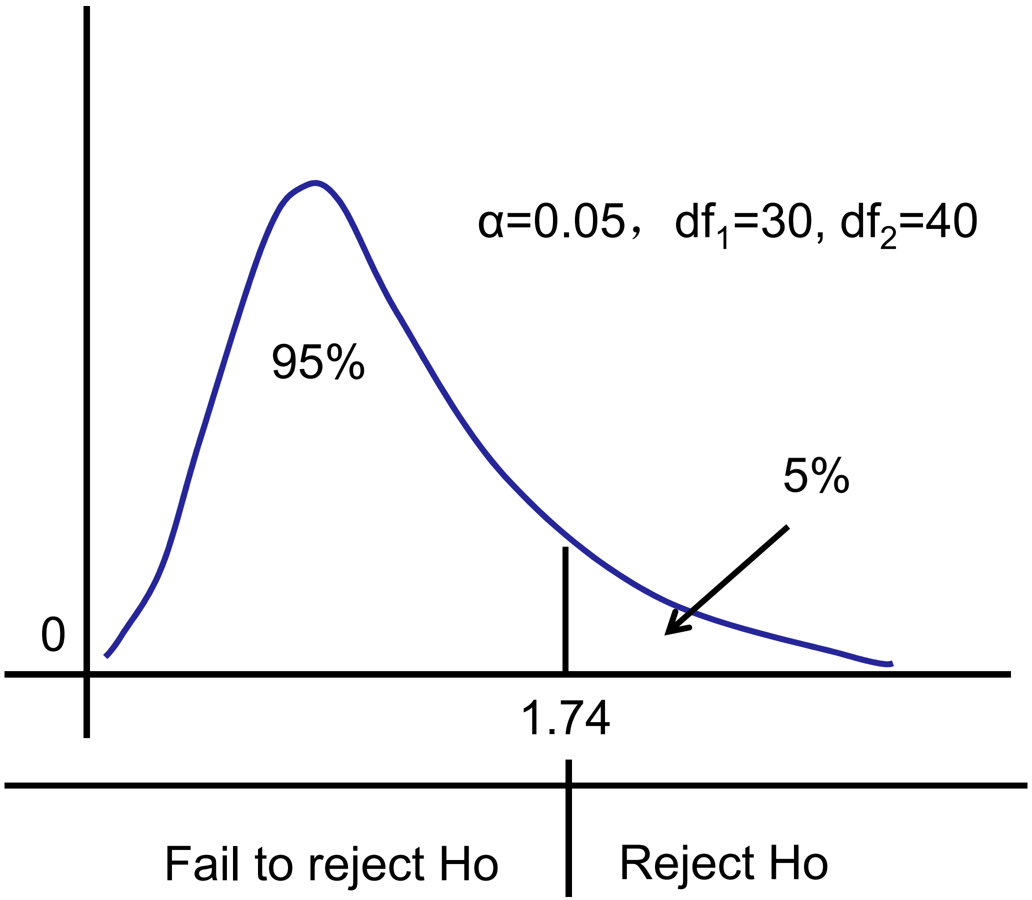
$$F = S_1^2 / S_2^2$$

- Specify the level of significance. Let's conduct our hypothesis test at the 5% level of significance.
- State the decision rule regarding the hypothesis. Using the sample sizes for the two industries, the critical F-value for our test is found to be 1.74. This value is obtained from the table of the F-distribution at the 5% level of significance with df1= 30 and df2 = 40. Thus, if the computed F-statistic is greater than the critical value of 1.74, the null hypothesis is rejected. The decision rule, illustrated in the figure below, can be stated as:

$$\text{Reject Ho if } F > 1.74$$

R11 Example: Hypothesis Testing

- Decision rule for F-test



R11 Example: Hypothesis Testing

- *Collect the sample and calculate the sample statistics*. Using the information provided, the F-statistic can be computed as:

$$F = \frac{S_1^2}{S_2^2} = \frac{\$4.30^2}{\$3.80^2} = \frac{\$18.49}{\$14.44} = 1.2805$$

- *Make a decision regarding the hypothesis*. Since the calculated F-statistic of 1.2805 is less than the critical F-statistic of 1.74, we fail to reject the null hypothesis.
- *Make a decision based on the results of the test*. Based on the results of the hypothesis test, Cower should conclude that the earnings variances of the industries are not statistically significantly different from one another at a 5% level of significance. More pointedly, the earnings of the textile industry am not more divergent than those of the paper industry.

R11 Example: Hypothesis Testing

- Which type of test is used to test if the square deviations of the two normal distribution population are equal?
  A. T-test
  B. χ^2-test
  C. F-test
- **Correct answer: C**
- William Adams wants to test whether the mean monthly returns over the last five years are the same for two stocks. If he assumes that the returns distributions are normal and have equal variances, the type of test and test statistic are best described as:
  A. Paired comparisons test, t-statistic
  B. Paired comparisons test, F-statistic
  C. Difference in means test, t-statistic
- **Correct answer: A**

R11 Hypothesis Testing

- Parametric tests
  - rely on assumptions regarding the distribution of the population
  - specific to population parameters.
  - For example, z-test.
- Nonparametric tests
  - either do not consider a particular population parameter or have few assumptions about the population that is sampled.
  - Nonparametric tests are used:
    - ✓ The assumptions about the distribution of the random variable that support a parametric test are not met.
      - Example: hypothesis test of the mean value for a variable that comes from a distribution that is not normal and is of small size so that neither the t-test nor the z-test are appropriate.
    - ✓ When data are ranks (an ordinal measurement scale) rather than values.
    - ✓ The hypothesis does not involve the parameters of the distribution, such as testing whether a variable is normally distributed.

R12 Technical Analysis

- Technical Analysis
  - the principles of technical analysis, its applications, and its underlying assumptions
  - Types of charts
  - the uses of trend
  - Common chart patterns
  - Common analysis indicators
  - the use of cycles

R12 Technical Analysis

- **Principles:**
  - Prices are determined by the interaction of supply and demand.
  - Only participants who actually trade affect prices, and better-informed participants tend to trade in greater volume.
  - Price and volume reflect the collective behavior of buyers and sellers.
- **Assumptions:**
  - Market prices reflect both rational and irrational investor behavior.
    - ✓ Investor behavior is reflected in trends and patterns that trend to repeat and can be identified and used for forecasting prices.
    - ✓ Efficient markets hypothesis dose not hold.

R12 Example: Technical Analysis

- 1. Technical analysis relies most importantly on:
  - A. price and volume data.
  - B. accurate financial statements.
  - C. fundamental analysis to confirm conclusions.
- 2. Which of the following is not an assumption of technical analysis?
  - A. Security markets are efficient.
  - B. The security under analysis is freely traded.
  - C. Market trends and patterns tend to repeat themselves.

R12 Technical Analysis

- The differences among technicians, fundamentalists and Efficient market followers.
  - Fundamental analysis of a firm attempts to determine the intrinsic value of an asset by using the financial statements and other information.
  - Technical analysis uses only the firm's share price and trading volume data, and it is not concerned with identifying buyers' and sellers' reasons for trading, but only with the trades that have occurred.
  - Fundamentalists believe that prices react quickly to changing stock values, while technicians believe that the reaction is slow. Technicians look for changes in supply and demand, while fundamentalists look for changes in value.

R12 Technical Analysis

- **Advantages of technical analysis:**
  - Actual price and volume data are observable.
  - Technical analysis itself is objective (although require subjective judgment), while much of the data used in fundamental analysis is subject to assumptions or restatements.
  - It can be applied to the prices of assets that do not produce future cash flows, such as commodities.
  - It can also be useful when financial statement fraud occurs.
- **Disadvantage:**
  - The usefulness is limited in markets where price and volume data might not truly reflect supply and demand, such as in <u>illiquid markets</u> and in markets that are subject to outside manipulation.

R12 Example: Technical Analysis

- Why is technical analysis especially useful in the analysis of commodities and currencies?
  - A. Valuation models cannot be used to determine fundamental intrinsic value for these securities.
  - B. Government regulators are more likely to intervene in these markets.
  - C. These types of securities display clearer trends than equities and bonds do.

R12 Technical Analysis

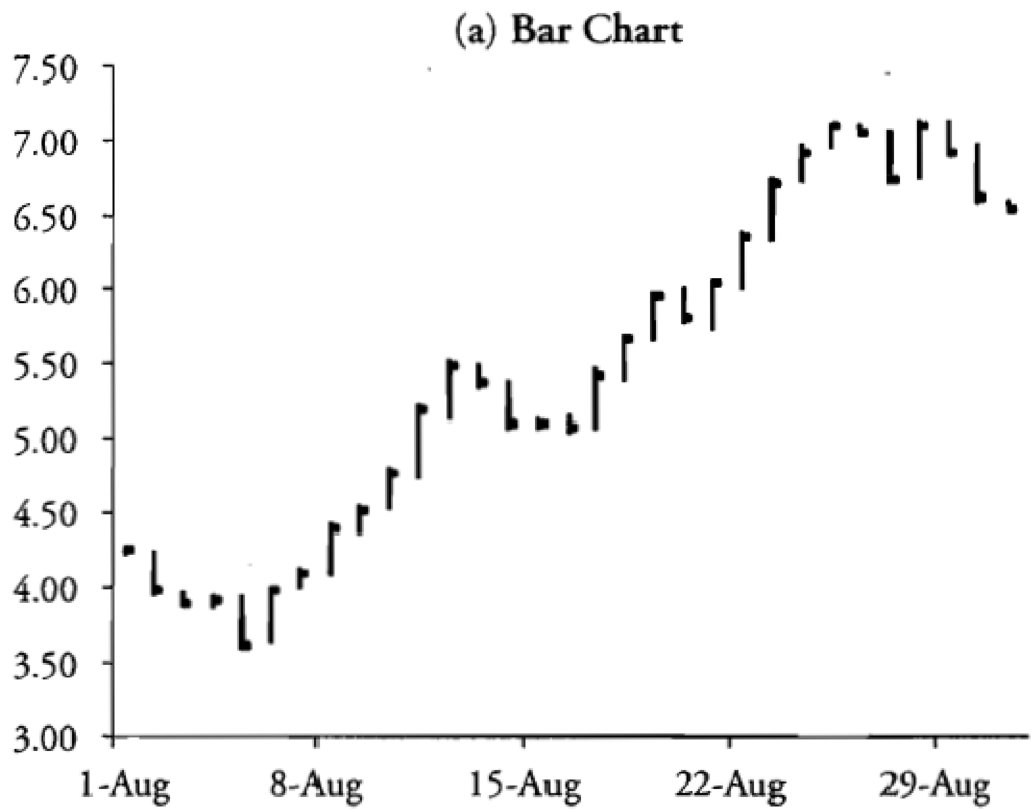
- Charts of price and volume are used to analyze asset prices and overall market movement.
  - Horizontal axis: usually time interval (daily, weekly, monthly)
  - Vertical axis: Price
- Types of charts:
  - Line charts
  - Bar charts
  - Candlestick charts
  - Point and figure charts

R12 Technical Analysis

Figure 1: Line Chart

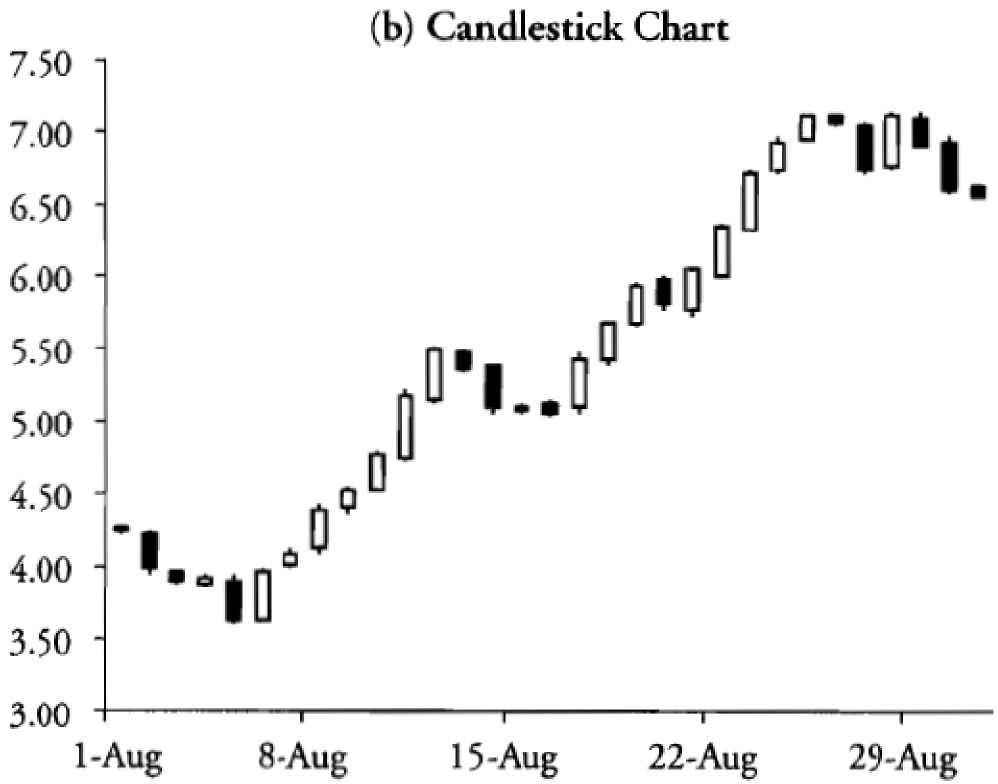
➢**Line Charts** are the simplest technical analysis charts. They show closing prices for each periods as a continuous line.

R12 Technical Analysis



➢**Bar charts** add the high and low prices for each trading period and often include the opening price and closing price as well.

R12 Technical Analysis



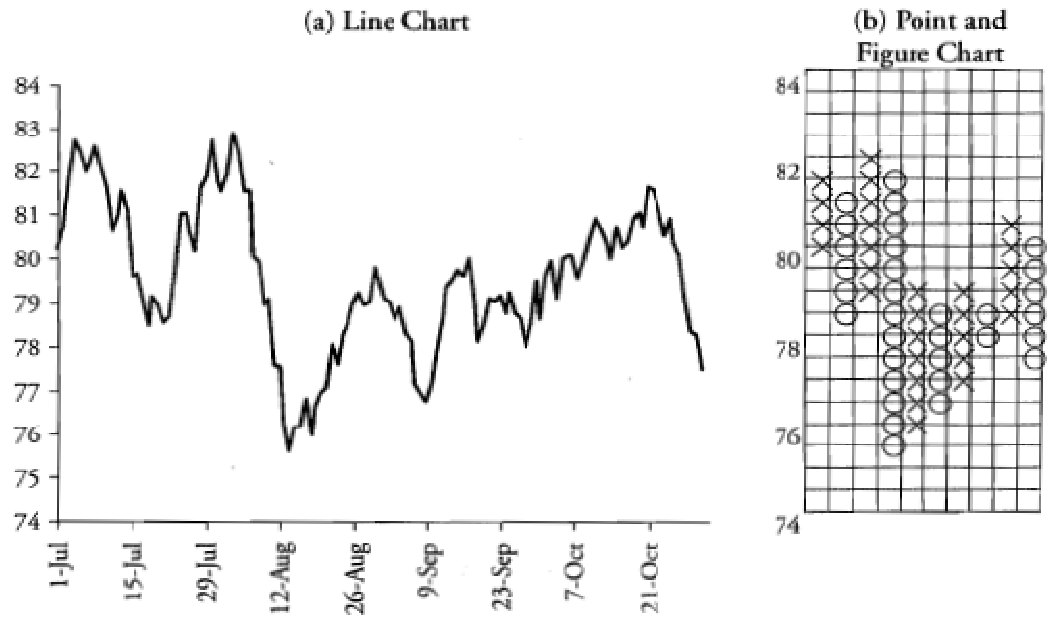
- **Candlestick charts** use the same data as bar charts but display a box bounded by the opening and closing prices.
  - •Box is clear: closing price>opening price;
  - •Box is filled: closing price<opening price

R12 Example: Technical Analysis

- A candlestick chart is similar to a bar chart except that the candlestick chart:
  A. represents upward movements in price with X's.
  B. also graphically shows the range of the period's highs and lows.
  C. has a body that is light or dark depending on whether the security closed higher or lower than its open.

R12 Technical Analysis

Figure 3: Charts of Price Data



➢**Point and figure charts** are helpful in identifying changes in the direction of price movements.

- •Starting form opening price;
- •X: increase of one box size, O: indicate a decrease.
- •Analyst will begin the next column when the price changes in the opposite direction by at least the reversal size (3 times the box size).

R12 Technical Analysis

- Relative strength analysis: an analyst calculate the ratios of an asset's closing prices to benchmark values, such as stock index or comparable asset, and draws a line chart of the ratios.
  - Positive relative strength: an increasing trend indicates that the asset is outperforming the benchmark
  - Negative relative strength: an decreasing trend indicates that the asset is underperforming the benchmark

R12 Technical Analysis

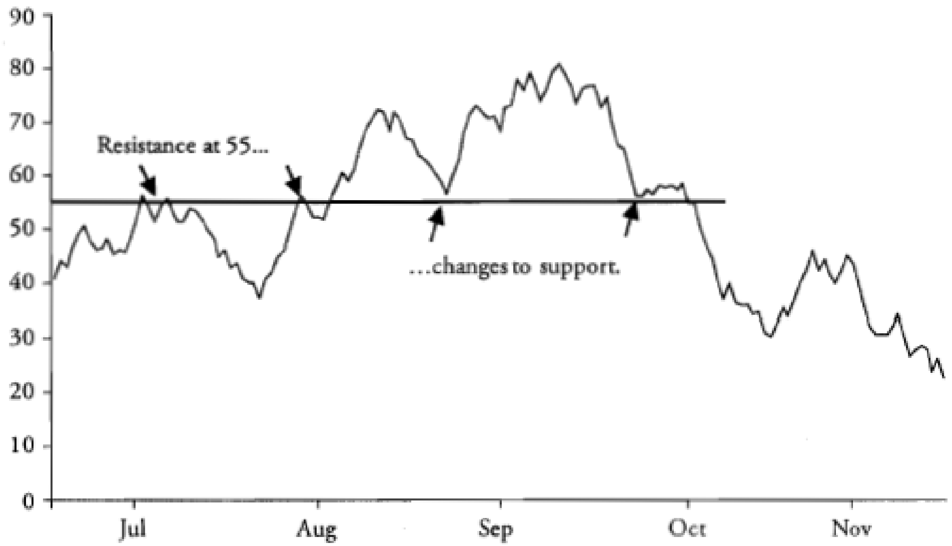
- **Trend**: is the most basic concept in technical analysis.
  - Uptrend: prices are consistently reaching higher highs and retracting to higher lows. (Demand>Supply)
  - Downtrend: prices are consistently reaching higher lows and retracting to lower highs. (Demand<Supply)
- **Trend line**: can help to identify whether a trend is continuing or revering.
  - Uptrend line: connects the increasing lows in prices;
  - Downtrend line: connects the decreasing highs in prices;
- When prices crosses the trend line by what the analyst considers a significant amount, a **breakout** form a downtrend or a **breakdow**n form an uptrend is said to occur.

R12 Example: Technical Analysis

- A downtrend line is constructed by drawing a line connecting:
  - A. the lows of the price chart.
  - B. the highs of the price chart.
  - C. the highest high to the lowest low of the price chart.

R12 Technical Analysis

- **Support level**: buying is expected to emerge that prevents further price decreases.
- **Resistance level:** selling is expected to emerge that prevents further price increases.
- **Change in polarity**: breached resistance levels become support levels and that breached support levels become resistance levels.



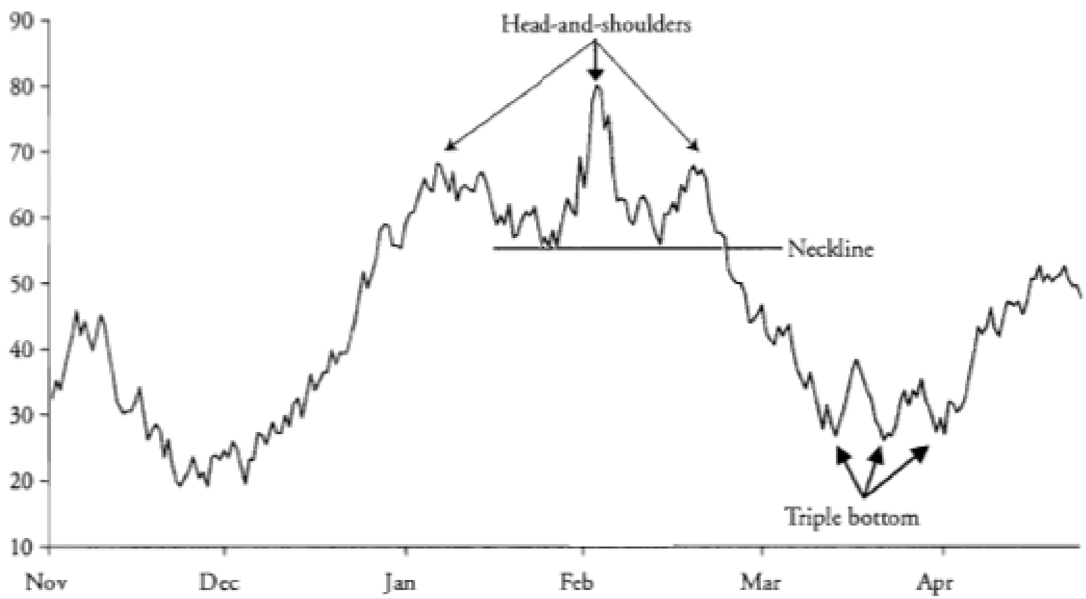
R12 Technical Analysis

➢ Common chart patterns.

- Reversal patterns
  - ✓ For uptrend: Head-and shoulders pattern, Double top and triple top
  - ✓ For downtrend: inverse head-and shoulders pattern, Double bottom, and triple bottom
- Continuation patterns
  - ✓ Triangles
  - ✓ Rectangles

R12 Technical Analysis

Figure 6: Reversal Patterns



➢**Head-and-shoulders pattern** is used to project a price target for ensuing downtrend.

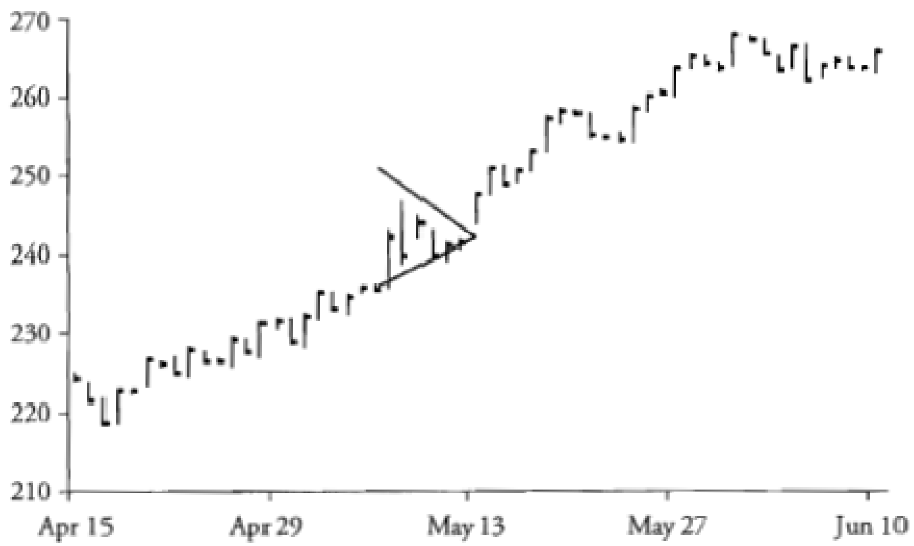
➢**The size of the head-and-shoulders pattern**: the difference in price between the **head** and the **neckline**.

➢**Price target = Neckline – (Head – Neckline)**

➢**Inverse head and shoulders pattern: price target = neckline + (neckline – head)**

R12 Technical Analysis

Figure 7: Triangle Continuation Pattern



- **Triangles:** form when prices reach lower highs and higher lows over a period of time.
- **Rectangles**: form when trading temporarily forms a range between a support level and a resistance level.
- **Flags and pennants**: refer to rectangles and triangles that appear on short-term price charts.

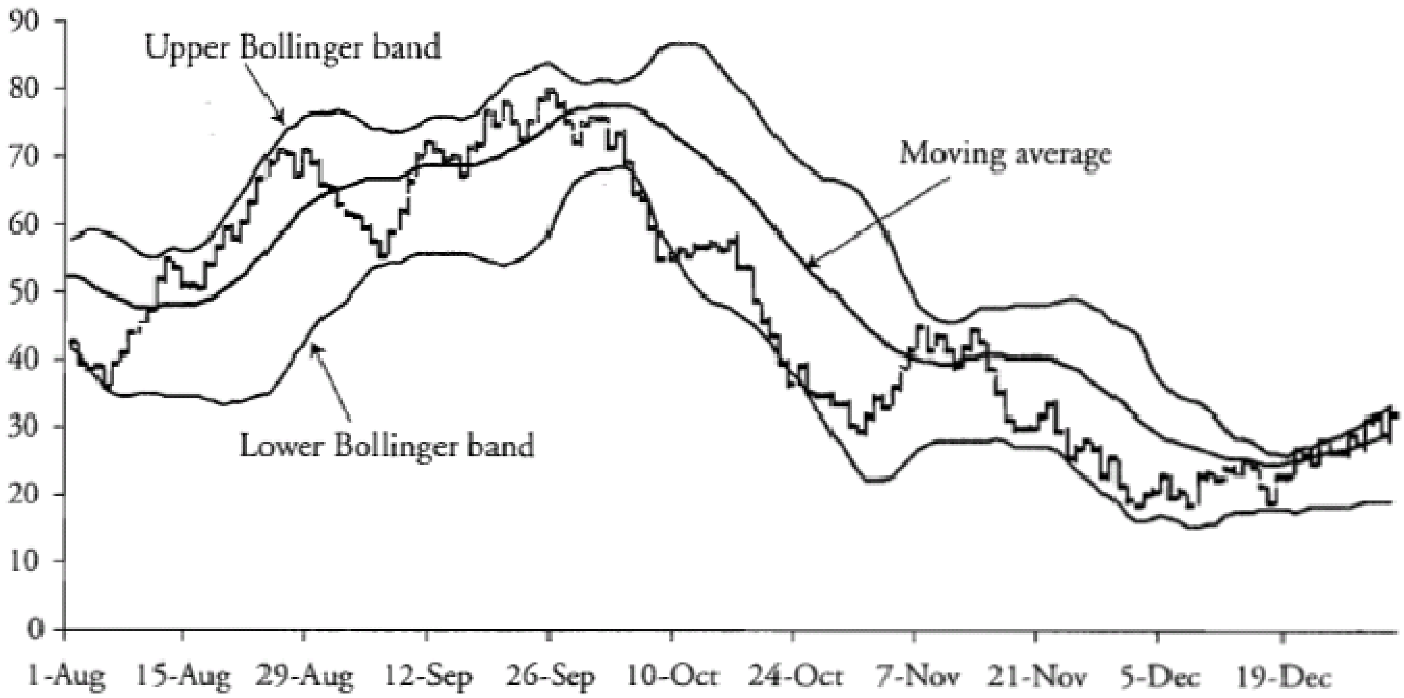
R12 Technical Analysis

- **Common analysis indicators**
  - Price-based
    - ✓ Moving average lines
    - ✓ Bollinger bands
  - Momentum oscillators
    - ✓ Rate of change oscillator
    - ✓ Relative Strength Index
    - ✓ Moving average convergence/divergence
    - ✓ Stochastic oscillator
  - Sentiment
    - ✓ Put/call ratio
    - ✓ Volatility Index
    - ✓ Margin debt
    - ✓ Short interest ratio
  - Flow of funds
    - ✓ Short-term trading index
    - ✓ Margin debt
    - ✓ Mutual fund cash position
    - ✓ New equity issuance

R12 Technical analysis indicators: price-based

- Moving average lines (移动平均线)
- Bollinger bonds(布林带或布林线)
  - Moving average +/- 2σ

Figure 8: Moving Average and Bollinger Bands



R12 Technical analysis indicators: Oscillators

- **Rate of Change Oscillator (ROC)**

$M = (V - Vx) \times 100$ or $M = \dfrac{V}{V_x} \times 100$

where

M = momentum oscillator value

V = last closing price

Vx = closing price x days ago, typically 10 days

- One way technical analysts use the ROC is to **buy** when the oscillates changes from negative to positive during a uptrend in prices, and **sell** when the ROC changes from positive to negative during a downtrend.

R12 Technical analysis indicators: Oscillators

➤ **Relative Strength Index (RSI)**

$$RSI = 100 - \frac{100}{1+RS}$$

$$where\ RS = \frac{\sum(\text{Up changes for the period under consideration})}{\sum(\text{Down changes for the period under consideration})}$$

- An RSI is based on the ratio of total price increases to total price decreases over a selected number of periods. This ratio is then scaled to oscillated between 0 and 100, with high values (typically those greater than 70) indicating an **overbought** market and low values (typically those less than 30) indicating an **oversold** market.

R12 Technical analysis indicators: Oscillators

- **Stochastic Oscillator**

$$\%K = 100\left(\frac{C - L14}{H14 - L14}\right)$$

where

C = latest closing price

$L14$ = lowest price in past 14 days

$H14$ = highest price in past 14 days

and

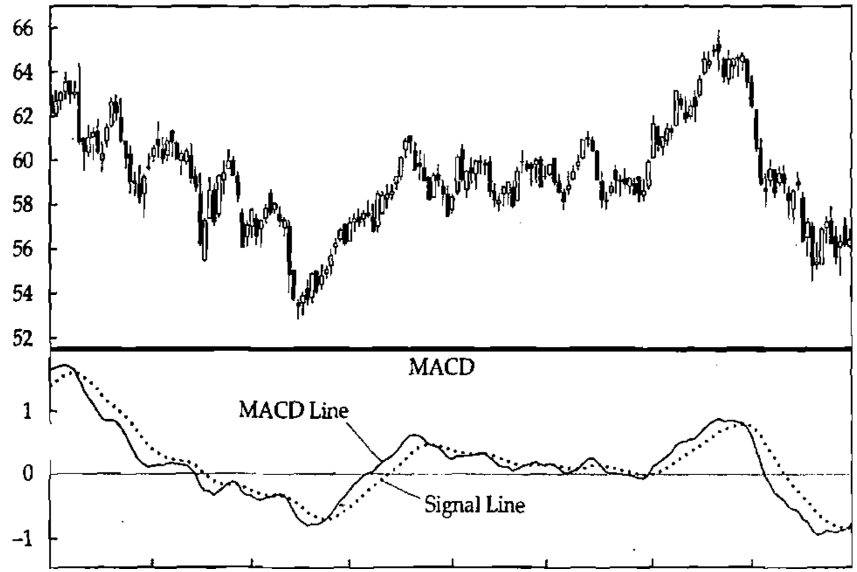
$\%D$ = average of the last three %K values calculated daily

- Technical analysts typically use stochastic oscillators to identity overbought and oversold markets. Points there the %D line crosses the %K line can also be used as trading signals.

R12 Technical analysis indicators: Oscillators

➢ Moving Average Convergence/Divergence (MACD)

- The MACD is the difference between a short-term and a long-term moving average of the security's price. The MACD is constructed by calculating two lines, the MACD line and the signal line:
- **MACD line**: difference between two exponentially smoothed moving averages, generally 12 and 26 days.
- **Signal line**: exponentially smoothed average of MACD line, generally 9 days.
  - ✓ *The MACD line crossing above the smoother signal line is viewed as a buy signal and the MACD line crossing below the signal line viewed as a sell signal.*

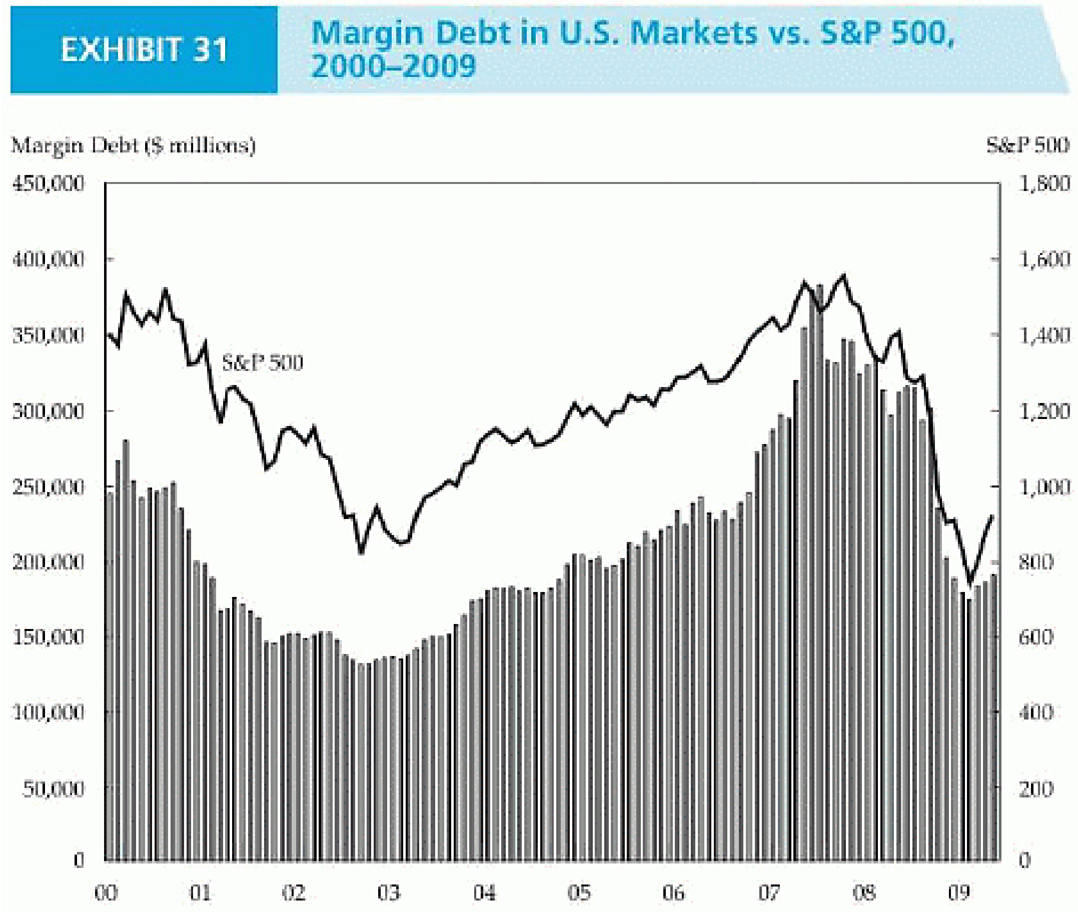


R12 Technical analysis indicators: Non-price-based

- **Put/call ratio**
  - The put /call ratio is put option volume divided by call option volume. Increases in the put/call ratio indicate a more negative outlook for the price of the asset. This ratio is generally viewed as a *contrarian indicator*. Extremely high ratios indicate strongly bearish investor sentiment and possibly an oversold market, while extremely low ratios indicate strongly bullish sentiment and perhaps an overbought market.
- **Volatility index (VIX)**
  - The Chicago Board Options Exchange calculates the VIX, which measures the volatility of options on the S&P 500 stock index. High levels of the VIX suggest investors fear declines in the stock market. *Technical analysts most often interpret the VIX in a contrarian way, viewing a predominantly bearish investor outlook as a bullish sign*.

R12 Technical analysis indicators: Non-price-based

- **Margin debt**
  - Increasing margin debt tends to coincide with increasing market prices and decreasing margin debt tends to coincide with decreasing prices.



100% Contribution Breeds Professionalism

R12 Technical analysis indicators: Non-price-based

- **Short interest ratio**
  - Investors sell shares short when they believe the share prices will decline. The number of shares of a particular security that are currently sold short is called "short interest."
  - *Short interest ratio = Short interest / Average daily trading volume*
  - *A high short interest ratio means investors expect the stock price to decrease, it also implies future buying when short sellers must return their borrowed shares*. Thus, technical analysts' opinions are divided as to how the short interest ratio should be interpreted.

R12 Technical analysis indicators: Non-price-based

Arms index or Short-term trading index (TRIN)

- The TRIN is a measure of funds flowing into advancing and declining stocks. The index is calculated as:

$$\text{Arms Index} = \frac{\text{Number of advancing issues} \div \text{Number of declining issues}}{\text{Volume of advancing issues} \div \text{Volume of declining issues}}$$

- And index value close to 1 suggests funds are flowing about evenly to advancing and declining stocks. Index values greater than 1 mean the majority of volume is in declining stocks, while an index less than 1 means more of the volume is in advancing stocks.

R12 Technical analysis indicators: Non-price-based

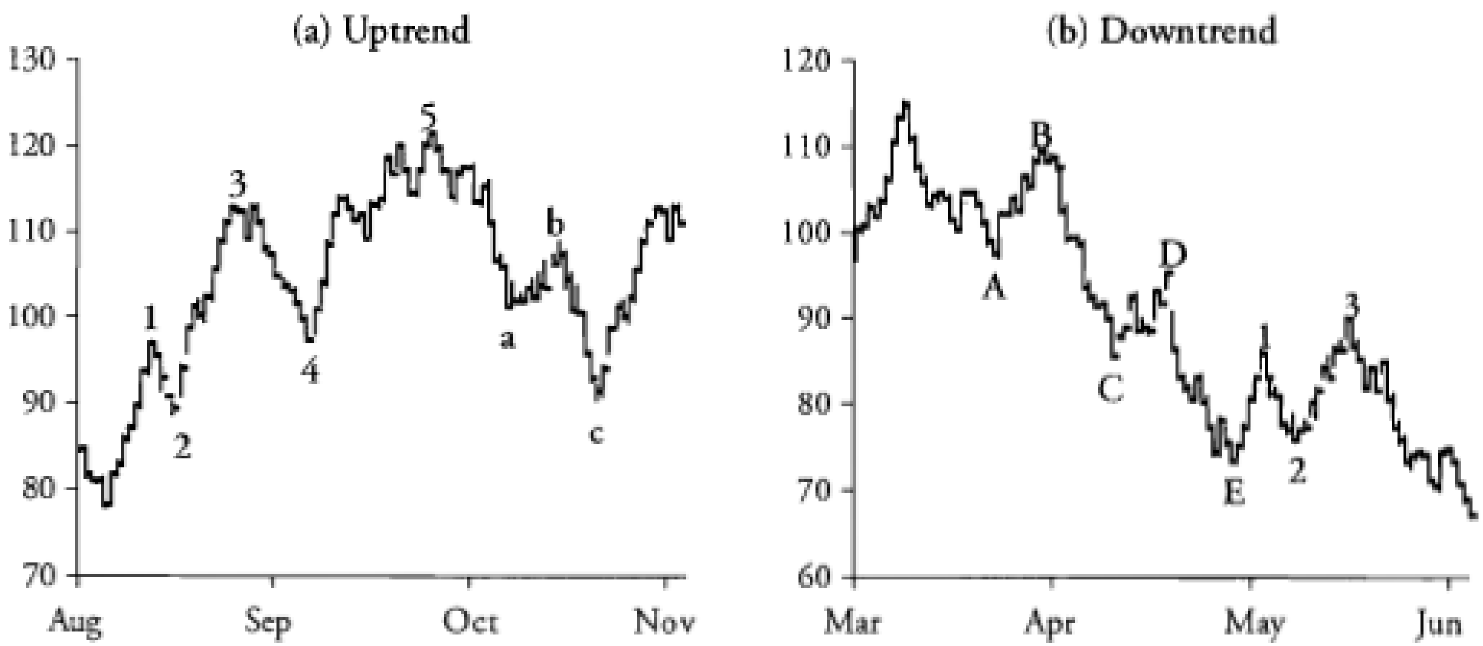
- **Mutual fund cash position**
  - Technical analysts typically view mutual fund cash as a *contrarian indicator*. When mutual funds accumulate cash, this represents future buying power in the market. A high mutual fund cash ratio therefore suggests market prices are likely to increase.
- **New equity issuance**
  - Issuers tend to sell new shares when stock prices are thought to be high, increases in issuance of new shares may often coincide with market peaks.

R12 Technical Analysis

- **Cycle theory**: is the study of processes that occur in cycles.
  - 4-year presidential cycles: related to election years in the USA
  - Decennial patterns: 10-year cycles
  - Kondratieff wave: 18-year cycles, 54-year cycles
- **Elliott wave theory**: is based on the belief that financial market prices can be described by an interconnected sets of cycles.
  - **Waves**: refer to chart patterns associated with Elliott wave theory.
  - **Fibonacci ratios**: the sizes of these waves are thought to correspond with Fibonacci ratios (0,1,1,2,3,5,8,13,21, and so on)

R12 Technical Analysis

Figure 9: Elliott Wave Patterns



- **Prevailing up trend**: upward moves in prices consist of 5 waves and downward moves occur in 3 waves
- **Prevailing down trend**: downward moves in prices consist of 5 waves and upward moves occur in 3 waves

R12 Technical Analysis

- **Intermarket analysis**: refers to analysis of the interrelationships among the market values of major asset classes, such as stocks, bonds, commodities and currencies.
- The approach is also useful for comparing the relative performance of equity market sectors or industries and of various international market.

It’s not the end but just the beginning.

Life is short. If there was ever a moment to follow your passion and do something that matters to you, that moment is now.

生命苦短，如果你有一个机会跟随自己的激情去做你认为重要的事，那么这个机会就是现在。