Kurtosis

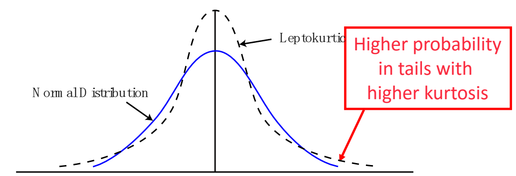
Kurtosis (Cont.)

- **Mesokurtic**
  - ✓ Identical to normal distribution;
  - ✓ Kurtosis = 3, excess kurtosis = 0.
- **Platykurtic**
  - ✓ Less peaked, thinner tailed than normal distribution;
  - ✓ Kurtosis < 3, excess kurtosis < 0.

Kurtosis

Kurtosis (Cont.)

➢ Leptokurtic vs. Normal distribution



Summary

➢ **Importance:** ☆☆

➢ **Content:**

- Skewness: positively & negatively, and the relative location of mean, median, and mode;
- Kurtosis: leptokurtic, mesokurtic, platykurtic.

➢ **Exam tips:**

- 常考点：正偏、负偏时mean、median、mode的相对位置。

Basic Concepts & Types of Probability

Tasks:

➢ Define basic concepts of probability, including random variable, outcome, and event;

➢ Distinguish different types of probability.

Probability

Terminology of probability

- Random variable
  - A quantity whose possible values are uncertain.
- Outcomes
  - The possible values of a random variable.
- Event
  - A specified set of outcomes.

Probability

Relationship among events

- **Mutually exclusive events**
  - The events competing directly with each other that they would not happen at same time.
- **Exhaustive events**
  - Events cover all possible outcomes.

Probability

Relationship among events (Cont.)

- **Independent events**
  - Occurrence of event A does not influence the occurrence of the event B, then A and B are independent.
- **Dependent events**
  - Occurrence of event A does influence the occurrence of the event B, then A and B are dependent.

Probability

Two defining properties of a probability

- The probability for any event E is: $0 \le P(E) \le 1$;
- For a set of events that are mutually exclusive and exhaustive, the sum of probabilities is 1: $\Sigma P(E_i) = 1$.

Probability

Types of probability

- **Empirical probability**
  - Estimate the probability of an event based on historical data.
- **Priori probability**
  - Deduced based on logical analysis rather than on observation or personal judgment.
- **Subjective probability**
  - Drawing on personal or subjective judgment without reference to any particular data.

Probability

Odds

- **Odds for the event E:** P(E)/[1-P(E)], given a probability P(E);
- **Odds against the event E:** [1-P(E)]/P(E), given a probability P(E).

Probability

Odds (Cont.)

- **Example:**

  Given P(horse will win the race)=1/8, what are the odds for or against the horse will win the race?

Answer:

Odds for horse will win the race = (1/8)/(1-1/8)=1/7;

Odds against horse will win the race = (1-1/8)/(1/8)=7/1.

Probability

Unconditional & conditional probability

- **Unconditional probability** (marginal probability): probability of an event (A) not conditioned on another event, denoted P(A).
  - E.g., probability of market will be up for the day.
- **Conditional probability:** probability of an event (A) conditioned on another event (B), denoted P(A｜B).
  - E.g., probability that the market will be up for the day, given that the Fed raises interest rates.

Summary

- **Importance:** ☆☆
- **Content:**
  - Relationship of events;
  - Types of probability: empirical, priori, subjective probability; odds for and odds against; conditional and unconditional probability.
- **Exam tips:**
  - 常考点： empirical, priori, subjective 概率的辨识。

Algorithms of Probability

Tasks:

- Explain multiplication, addition, and total probability rules and master their application;
- Calculate and interpret updated probability using Bayes' formula.

Probability

Joint probability and multiplication rule

- **Joint probability:** the probability of event A and B both happen, denoted P(AB).
- **Multiplication rule for joint probability:**
  - P(AB)=P(A | B)P(B)

    ✓P(A | B)=P(AB)/P(B)
  - For independent events: P(A | B)=P(A) → P(AB)=P(A)P(B).

Probability

Example

- If the probability that interest will increase P(I) is 40%, and if the interest increase, the probability that economy will go into a recession P(R|I) is 70%. What is the probability of a recession **and** an increase in interest rates P(RI)?

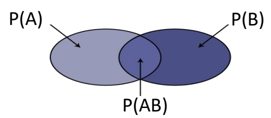
Answer:

P(RI) = P(R|I) × P(I) = 0.7 × 0.4 = 28%.

Probability

Probability that at least one of two events occur

- Denoted as P(A or B), or P(A+B)
- P(A or B) = P(A occur and B doesn't occur)
  + P(A doesn't occur and B occur)
  + P(A and B both occur)
  - **P(A + B)= P(A) + P(B) - P(AB)**



Don't double count P(AB)

Probability

Probability that at least one of two events occur (Cont.)

- If event A and B are mutually exclusive, then:
  - P(AB)=0 ⟶ **P(A + B)= P(A)+P(B)**

Example: If the probability that interest rate will increase P(I) is 40%, the probability that economy will go into a recession P(R) is 34%, and the joint probability P(RI) is 28%. What is the probability of either interest rates increase **or** recession?

Answer:

P(R + I) = P(R) + P(I) – P(RI) = 34% + 40% – 28% = 46%.

Probability

Total probability rule

- Explains the unconditional probability of the event in terms of probabilities conditional on the scenarios.
  - $P(A) = P(A|S_1)P(S_1) + P(A|S_2)P(S_2) + \ldots\ldots + P(A|S_n)P(S_n)$

    Wherein: S_1, S_2, … … S_n are mutually exclusive and exhaustive.

Probability

Example

- If the probability that interest rate will increase P(I) is 0.4 and the probability that interest rate will not increase $P(I^c)$ is 0.6. Given that interest rate increase, the probability that economy will go into a recession is 0.7. Given that interest rate doesn't increase, the probability that economy will go into a recession is 0.1. What is the (unconditional) probability of recession P(R)?

Probability

Example (Cont.)

Answer:

$P(R) = P(R|I) \times P(I) + P(R|I^C) \times P(I^C)$

$= 0.70 \times 0.40 + 0.10 \times 0.60 = 0.34$

Probability

Bayes' formula

- Given a prior probabilities P(A) for an event of interest, if you receive new information (B), the rule for updating your probability (updated probability, P(A | B)) of the event.

$$P(A|B) = \frac{P(B|A)}{P(B)} \times P(A)$$

Probability

Example

- The lie detector can be used to detect if the suspect lies or not. It is known that the probability that the suspect lies is 0.7. If the suspect lies, the probability that the test result is "lied" is 0.9; if the suspect doesn't lie, the probability that the test result is "lied" is 0.2. What is the probability that the suspect does lie given the test result is "lied"?

Probability

Answer:

- Event A: the suspect lies, so $P(A) = 0.7$, $P(A^C) = 0.3$;
- Event B: the test result is "lied", so $P(B | A) = 0.9$, $P(B | A^C) = 0.2$. Using the total probability formula:
  - $P(B) = P(B | A) \times P(A) + P(B | A^C) \times P(A^C)$
    $= 0.9 \times 0.7 + 0.2 \times 0.3 = 0.69$
- Then, the probability that the suspect does lie given the test result is "lied" is:

$$P(A|B) = \frac{P(B|A)}{P(B)} \times P(A) = \frac{0.9}{0.69} \times 0.7 = 0.913$$

Summary

- **Importance:** ☆☆☆
- **Content:**
  - Multiplication, addition, and total probability rules;
  - Bayes' formula.
- **Exam tips:**
  - 常考点：考计算题。

Application of Probability in Investment

Tasks:

- Calculate and interpret expected value, variance, standard deviation, and covariance and correlation of returns on a portfolio;
- Identify the most appropriate method to solve a particular counting problem.

Probability

Probability weighted expected value E(X)

- The probability-weighted average of the possible outcomes of the random variable (X).

  $E(X) = P(x_1)x_1 + P(x_2)x_2 + \dots + P(x_n)x_n$

- Total probability rules for expected value:

  $E(X) = E(X|S_1)P(S_1) + E(X|S_2)P(S_2) + \dots\dots + E(X|S_n)P(S_n)$

  Wherein: S_1, S_2, … … S_n are mutually exclusive and exhaustive.

Probability

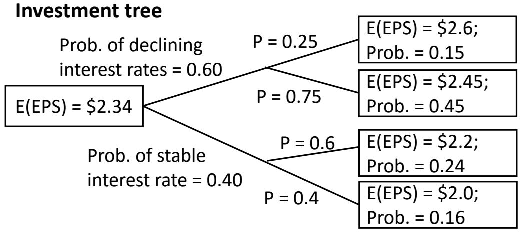
Probability weighted expected value (Cont.)

- An example of expected value about an economy:

	$P(x_i)$	Return (X_i)	$P(x_i)x_i$
Recession	0.25	-0.10	-0.025
Normal	0.50	0.08	0.040
Boom	0.25	0.22	0.055
			E(X) = 0.07

Probability

Investment tree



E(EPS) = 2.6 x 0.15 + 2.45 x 0.45 + 2.2 x 0.24 + 2.0 x 0.16

= $2.34

Probability

Probability weighted variance and standard deviation

- Variance

  $\sigma^2(X)=P(x_1)[x_1-E(X)]^2+P(x_2)[x_2-E(X)]^2+......+P(x_n)[x_n-E(X)]^2$
- Standard deviation: positive squared root of variance.

Example

	$P(x_i)$	$Return(x_i)$	$P(x_i)x_i$	$P(x_i)[x_i - E(X)]^2$
Recession	0.25	–0.10	–0.025	0.00723
Normal	0.50	0.08	0.040	0.00005
Boom	0.25	0.22	0.055	0.00563
			E(X) = 0.07	σ^2 =0.01290

Covariance and Correlation

Covariance

- A measure of how two variables move together.

  $$Cov(R_i,R_j)=\sum_{i,j=1}^{n}P(R_i,R_j)\left[R_i-E(R_i)\right]\left[R_j-E(R_j)\right]$$
  - Positive covariance: the two variables tend to be above or below their expected values at the same time;
  - Negative covariance: one variable tend to be above its expected value when the other is below its expected value.

Covariance and Correlation

Covariance (Cont.)

- Example of covariance given a joint probability function:

Returns	R_B = 40%	R_B = 20%	R_B = 0%	**$E(R_B)$ = 18%**
R_A = 20%	0.15			
R_A = 15%		0.60		
R_A = 4%			0.25	
$E(R_A)$ = 13%				

Cov_{AB} = 0.15(0.20 – 0.13)(0.40 – 0.18) + 0.6(0.15 – 0.13)(0.20 – 0.18) + 0.25(0.04 – 0.13)(0 – 0.18)

= 0.0066

Covariance and Correlation

Covariance (Cont.)

- Disadvantage of covariance:
  - Values range from minus infinity to positive infinity;
  - Units of covariance difficult to interpret.

Covariance and Correlation

Correlation

- A standardized measure of **linear** relationship between two variables.

$$\rho_{i,j}=\frac{Cov(R_i,R_j)}{\sigma_i\sigma_j},\ r_{i,j}=\frac{Cov(R_i,R_j)}{s_i s_j}$$

  - Values range from +1 (perfect positive correlation) to -1 (perfect negative correlation);
  - A correlation of 0 (uncorrelated variables) indicates an absence of any linear(straight-line) relationship;
  - The bigger the absolute value, the stronger the linear relationship.

Labeling, Combination, and Permutation

Labeling

- The number of ways that n objects can be labeled with k different labels, with n_1 of the first type,n_2 of the second type, and so on, with $n_1 + n_2 + ... + n_k = n$.

$$\text{Number of ways} = \frac{n!}{n_1!n_2!\ldots n_k!}$$

Labeling, Combination, and Permutation

Example

- Out of 10 stocks, 5 will be rated buy, 3 will be rated hold, and 2 will be rated sell. How many ways are there to do this?

Answer: $\frac{10!}{5!\times 3!\times 2!}=2,520$

Labeling, Combination, and Permutation

Combination

- The number of ways that we can choose r objects from a total of n objects, when the order in which the r objects are listed **does not matter.**

$$C_n^r = \frac{n!}{(n-r)!r!}$$

Labeling, Combination, and Permutation

Example

- Out of 10 stocks, 5 will be rated buy, the order of stock purchase does matter. How many ways are there to do this?

Answer: $\frac{10!}{(10-5)!5!} = 252$

Labeling, Combination, and Permutation

Permutation

- The number of ways that we can choose r objects from a total of n objects, when the order in which the r objects are listed **does matter.**

$$P_n^r = \frac{n!}{(n-r)!}$$

Labeling, Combination, and Permutation

Example

- Out of 10 stocks, 5 will be rated buy, the order of stock purchase does matter. How many ways are there to do this?

Answer: $\frac{10!}{(10-5)!} = 30240$

Summary

- Importance: ☆☆
- Content:
  - Probability weighted mean, variance and standard deviation;
  - Covariance and correlation;
  - Labeling, combination, and permutation.
- Exam tips:
  - 常考点：covariance and correlation的概念和计算。

Basics of Probability Distribution

Tasks:

- Define probability distributions;
- Interpret cumulative distribution function;
- Define discrete uniform random variable, Bernoulli random variable, and binomial random variable.

Probability Distribution

Probability distribution

- Specifies the probabilities of all possible outcomes for a random variable.
  - **Discrete distribution**: the distribution of the discrete random variable.
    - ✓ **Discrete random variable**: takes on a finite and countable number of possible values.

Probability Distribution

Probability distribution (Cont.)

- Specifies the probabilities of all possible outcomes for a random variable.
  - **Continuous distribution:** the distribution of the continuous random variable.
    - ✓ **Continuous random variable:** takes on an infinite and uncountable number of possible values.

Probability Distribution

Probability function

- Specifies the probability that the **discrete random variable** takes on **a specific value**.
  - P(X = x) is the probability that a random variable X takes on the value x .

Probability density function, f(x)

- Specifies the probability that the **continuous random variable** takes on a value **within a range**.
  - The probability of taking on an specific value is always zero, P(X=x)=0.

Probability Distribution

Cumulative probability function, F(x)

- Gives the probability that a random variable X is less than or equal to a particular value x, $P(X \le x)$.
  - For both discrete and continuous random variables:
    - ✓ $F(x) = P(X \le x)$;
    - ✓ $P(x_1 < X \le x_2) = F(x_2) - F(x_1)$.

Discrete Distribution

Discrete uniform distribution

- Has a finite number of possible outcomes, all of which are equally likely.

Example

- P(x) = 0.2, for X = {1,2,3,4,5}; then:
  - P(2) = 20%;
  - F(3) = 60%;
  - $P(2 \le X \le 4)$ = P(2)+P(3)+P(4) = 60%.

Discrete Distribution

Bernoulli random variable (Bernoulli trial)

- Random variables with only two outcomes, one represents success (denoted as 1), the other represents failure (denoted as 0).

Discrete Distribution

Binomial random variable

- The number of successes in **n** Bernoulli trials, assuming that:
  - The probability of success (p) is constant for all trials;
  - The trails are all independent.
- Expected value for binomial random variable = np;
- Variance for binomial random variable = np(1-p).

Discrete Distribution

The probability of binomial random variable

$$P(x) = C_n^x p^x (1-p)^{n-x} = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$$

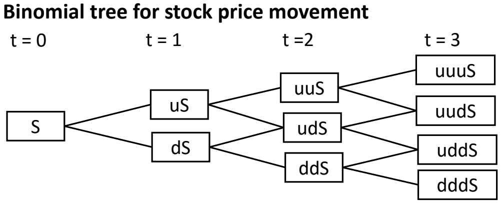
Example

- What is the probability of drawing exactly two white marbles from a bowl of white and black marbles in six tries if the probability of selecting white is 0.4 each time?

Answer: $P(2)=\frac{6!}{(6-2)!2!}(0.4)^2(1-0.4)^{6-2}=0.31$

Discrete Distribution

Binomial tree for stock price movement



- Up factor (u) > 1; down factor (d) = 1/u.
- Probability of move-up = p, probability of move-down = 1-p; E.g., $P(uudS)=\frac{3!}{(3-2)!2!}p^2(1-p)$

Discrete Distribution

Tracking error

- The total return on the portfolio (gross of fees) minus the total return on the benchmark index.
  - Tracking risk: the standard deviation of tracking error;
  - Tracking risk can be used to assess the manager’s performance.

Summary

- **Importance:** ☆☆
- **Content:**
  - Discrete and continuous distribution;
  - Probability function and probability density function;
  - Discrete uniform distribution and binomial distribution.
- **Exam tips:**
  - 常考点：binomial distribution 下的概率计算。

高顿网校 www.gaodun.com

Continuous Probability Distributions (1)

Tasks:

- Define continuous uniform distribution;
- Explain key properties of the normal distribution and master its application.

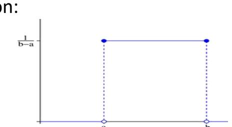
高顿网校 www.gaodun.com

Continuous Distribution

Continuous uniform distribution

- Probability of continuous uniform random variable which distribute evenly over an interval.
  - Probability density function:

$$f(x)=\begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{otherwise} \end{cases}$$



  - Probability for an interval:

$$P(a \le x \le b) = \int_a^b f(x)\,dx$$

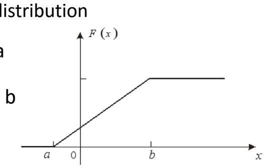
高顿网校 www.gaodun.com

Continuous Distribution

Continuous uniform distribution (Cont.)

- Cumulative probability distribution

$$F(x)=\begin{cases} 0 & \text{for } x \le a \\ \frac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } x \ge b \end{cases}$$



Continuous Distribution

Normal distribution

- Properties of normal distribution:
  - Completely described by mean and variance;
  - Skewness = 0 (symmetric about the mean);
  - Kurtosis = 3
  - Linear combination of normally distributed random variables is also normally distributed;
  - Probabilities decrease further from the mean, but the tails go on forever.

Continuous Distribution

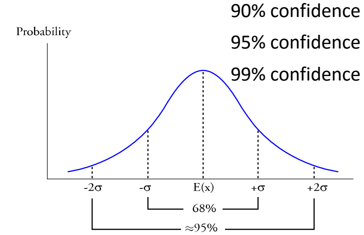
Normal distribution (Cont.)

- Probability for a given interval

90% confidence interval = μ +/- 1.65s

95% confidence interval = μ +/- 1.96s

99% confidence interval = μ +/- 2.58s



Continuous Distribution

Standard normal distribution (z-distribution)

- Normal distribution with mean $\mu = 0$, and standard deviation $\sigma = 1$.
- **Standardization**:

$$z = \frac{X - \mu}{\sigma}$$

  - The Z value calculated by standardization represents the number of the standard deviations from the mean.

Continuous Distribution

Example

- The EPS for a large group of firms are normally distributed and has μ = \$4.00 and σ = \$1.50. Find the probability that a randomly selected firm's earnings are less than \$3.70.

z	0	0.01
0.1	0.5298	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217

Continuous Distribution

Answer:

$$z=\frac{3.70-4.00}{1.50}=-0.20$$

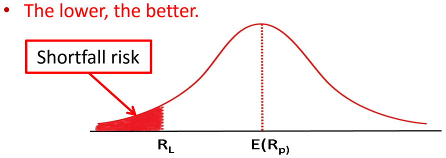
N(-0.2) = 1 – N(0.2) = 1 - 0.5793 = 0.4207 = **42.07%**

(3.70 is 0.20 standard deviations **below** the mean of 4.00)

Continuous Distribution

Normal distribution: application

- **Shortfall risk**: the risk that portfolio value or return will fall below the minimum acceptable level (R_L, shortfall level, threshold level) over some time horizon.
  - The lower, the better.



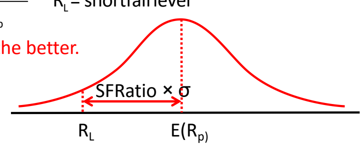
Continuous Distribution

Normal distribution: application (Cont.)

- **Safety-first ratio**: the distance from the mean return to the shortfall level in units of standard deviation.

$$SFRatio=\frac{E(R_p)-R_L}{\sigma_p} \qquad R_L = \text{shortfall level}$$

- The higher, the better.



Minimizing shortfall risk = Maximizing safety-first ratio

Summary

- **Importance:** ☆☆☆
- **Content:**
  - Continuous uniform distribution;
  - Properties of normal distribution and z-distribution;
  - Shortfall risk and safety first ratio.
- **Exam tips:**
  - 常考点1：normal distribution 的性质，考概念题；
  - 常考点2：shortfall risk and safety first ratio，考计算题和辨析题。

Continuous Probability Distributions (2)

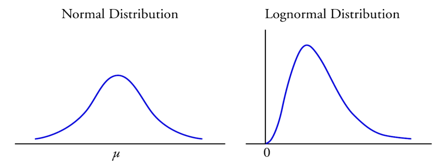
Tasks:

- Explain the relationship between normal and lognormal distributions;
- Describe properties of t-distribution;
- Compare Monte Carlo simulation and historical simulation.

Continuous Distribution

Lognormal distribution

- X is normally distributed, then e^x is lognormal distributed;
- Lognormal random variable is bounded from below by zero;
- It is positively skewed;
- It is used to model asset prices.



Continuous Distribution

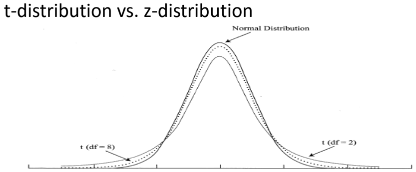
Student's t-distribution

- Defined by single parameter: degrees of freedom (df);
  - df = n – 1, where: n is the sample size.
- Symmetrical (bell shaped), skewness = 0;
- Fatter tails than a normal distribution;
- As df increase, t-distribution is approaching to standard normal distribution.

Continuous Distribution

Student's t-distribution

- t-distribution vs. z-distribution



- Given a degree of confidence, t-distribution has a wider confidence interval than z-distribution.

Continuous Distribution

Univariate distribution

- Describes a single random variable.

Multivariate distribution

- Describes the probabilities for a group of related random variables.
  - n means in total;
  - n variances in total;
  - n(n-1)/2 distinct correlations in total.

Discrete and Continuous Compounding

Discrete compounding

- $EAR = (1 + \text{Periodic interest rate})^m - 1 = (1 + \frac{r_s}{m})^m$

Continuous compounding

- $EAR = e^{r_s} - 1$
- $r_s = \ln(1 + EAR)$ or $r_s = \ln(1 + HPR_{1\text{-year}})$

Discrete and Continuous Compounding

Example

- If the continuously compounded stated rate = 8%, what is the effective holding period return for one and one-half years? How much will $1,200 grow to in one and one-half years?

Answer:

Effective holding period return = $e^{1.5\,(0.08)} - 1 = 12.75\%$

Amount would grow to in one-half years = $1,200 \times e^{1.5(0.08)}$

= $1,353.

Simulation

Monte Carlo simulation

- Use randomly generated values for risk factors, based on their assumed distributions, to produce a distribution of possible outcome.
- Limitation:
  - Fairly complex;
  - Do not directly provide precise insights;
  - Provide answer no better than the assumption used.

Simulation

Historical simulation

- Used randomly selected past data in risk factors to produce a distribution of possible outcome.
- Limitation:
  - Grounded in actual data and reflect only the risks represented in the sample historical data;
  - Does not lend itself to “what if” analyses.

Summary

- **Importance:** ☆☆
- **Content:**
  - Lognormal distribution and t-distribution;
  - Monte Carlo simulation vs. historical simulation.
- **Exam tips:**
  - 常考点1：t-distribution 的性质，考概念题；
  - 常考点2：两种simulation方法的优缺点比较，考概念题。

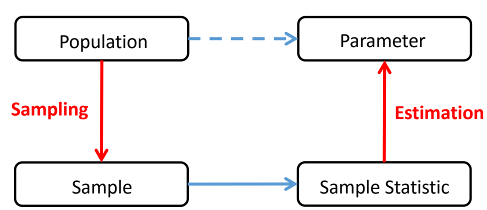
Sampling

Tasks:

- Distinguish between simple random and stratified random sampling;
- Define sampling error and sampling distribution;
- Describe the issues regarding sampling.

Sampling & Estimation

An overview



Sampling

Simple random sampling

- Each element of the population has an equal probability of being selected to the subset.

Stratified random sampling

- Step 1: separate the population into subpopulations based on one or more classification criteria;
- Step 2: use simple random sampling to draw from each stratum in sizes proportional to the relative size of each stratum in the population and then pooled to form a stratified random sample.

Sampling

Sampling error

- Difference between the sample statistic and the population parameter.

Sampling distribution

- The distribution of all the distinct possible values that the statistic can assume when computed from samples of the same size randomly drawn from the same population .

Sampling

Time-series data

- A sequence of returns collected at discrete and equally spaced intervals of time.
  - E.g., monthly prices for IBM stock for five years.

Cross-sectional data

- Data on some characteristic of individuals, groups, geographical regions, or companies at a single point in time.
  - E.g., returns on all health care stocks last month.

Sampling

Selection of sample size

- **Pros** for larger sample size
  - Larger sample size would produce a better estimate for parameter (better precision).
- **Cons** for larger sample size
  - Larger sample size may involve additional expenses that outweigh the value of additional precision;
  - Sampling from more than one population would not improve the estimate for the parameter (sampling risk).

Sampling

Sampling biases

- **Data-mining bias**: misuse of data referring to the repeatedly “drilling” in the same data until finding the statistically significant patterns.
- **Sample selection bias**: certain databases are excluded from the analysis.
  - **Survivorship bias**: only sampling from existing database.

Sampling

Sampling biases (Cont.)

- **Look-ahead bias**: using information that was not available on the test date.
- **Time-period bias**: based on a time period that may make the results time-period specific.

Summary

- **Importance:** ☆☆
- **Content:**
  - Simple random sampling vs. stratified random sampling;
  - Sampling error & sampling distribution;
  - Data-mining bias, sample selection bias (survivorship bias), look-ahead bias, time-period bias.
- **Exam tips:**
  - 常考点1：两个抽样方法的辨析；
  - 常考点2：data-mining bias和survivorship bias的辨析。

Point Estimate & Interval Estimate

Tasks:

- Identify and describe desirable properties of an estimator;
- Distinguish between point estimate and confidence interval estimate of a population parameter.

Estimation

Point estimate

- The calculated value of the sample statistic in a given sample, used as an estimate of the population parameter.
  - **Estimator**: formulas to compute the sample statistics;
    - ✓E.g., $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
    - ✓An estimator has a sampling distribution.
  - **Estimate**: particular value calculated from sample observations using an estimator.
    - ✓An estimate is a fixed number.

Estimation

Point estimate (Cont.)

- Desirable properties of an estimator:
  - **Unbiasedness**: the expected value equals the parameter it is intended to estimate;
  - **Efficiency**: the unbiased estimator of the population parameter that has a sampling distribution with smallest variance;

Estimation

Point estimate (Cont.)

- Desirable properties of an estimator:
  - **Consistency**: the probability of estimates close to the value of the population parameter increases as sample size increases.

Estimation

Confidence interval estimate

- Confidence interval for observation
  - A range for which a given percentage (1-α, called the **degree of confidence**) of all observations will lie based on a particular probability distribution.
    - ✓**Significance level (α):** the probability that the observations would not fall in a specific range.

Estimation

Confidence interval estimate (Cont.)

- Confidence interval

  = Point estimate ± Reliability factor x Standard deviation

  ✓Reliability factor = a number based on the assumed distribution of the population and the degree of confidence (1-α) for the confidence interval.

Estimation

Confidence interval estimate (Cont.)

- Reliability factors for normal distribution:
  - 90% confidence intervals: $Z_{0.05}$ = 1.65;
  - 95% confidence intervals: $Z_{0.025}$ = 1.96;
  - 99% confidence intervals: $Z_{0.005}$ = 2.58.

Estimation

Example

- The mean P/E (point estimate) for a sample of 41 firms is 19.0, sample standard deviation is 6.6. The population is normal distributed, calculated the 95% confidence interval for a randomly chosen firm.

Answer:

The reliability factor of 95% confidence level for normal distribution is 1.96.

Confidence interval: 19 +/- 1.96 x 6.6; or 6.06 to 31.94.

Summary

- **Importance:** ☆☆
- **Content:**
  - Point estimate and desirable properties;
  - Confidence interval estimate and reliability factors for normal distribution.
- **Exam tips:**
  - 正态分布不同置信区间对应的reliability factors，一定要记住。

Central limit theorem

Tasks:

- Explain central limit theorem；
- Calculate and interpret confidence interval for a population mean, given a normal distribution with known and unknown population variance, and unknown variance and a large sample size.

Estimation

Central limit theorem

- Given a population described by any probability distribution having mean μ and finite variance σ^2, the sampling distribution of the sample mean $\overline{X}$, computed from samples of size n from this population, will be approximately normal with mean μ (the population mean) and variance σ^2/n (the population variance divided by n) when the sample size n is large.

Estimation

Standard error

- The standard deviation of the distribution of sample statistic (sampling distribution).
  - **Standard error of sample mean:** the standard deviation of the distribution of sample means.
    - ✓ When the population standard deviation (σ) is **known**: $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$
    - ✓ When the population standard deviation (σ) is **unknown:** $s_{\overline{x}} = \frac{s}{\sqrt{n}}$

Estimation

Example

- The mean P/E for a sample of 41 firms is 19.0, and the standard deviation of the population is 6.6. What is the standard error of the sample mean?

Answer: $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6.6}{\sqrt{41}} = 1.03$

- Interpretation: for samples of size n = 41, the distribution of the sample means would have a mean of 19.0 and a standard deviation of 1.03.

Estimation

Confidence interval estimate (Cont.)

- Confidence interval for population mean
  - A range that contain the population mean with a given confidence level (1-α).
- The central limit theorem can be used to construct confidence intervals for population means:
  - Confidence interval of population mean
    = Point estimate of population mean
    ± Reliability factor x Standard error of sample mean

Estimation

Example

- The mean P/E (point estimate) for a sample of 41 firms is 19.0, sample standard deviation is 6.6. The population is normally distributed. What is the 95% confidence interval for the population mean?

Answer:

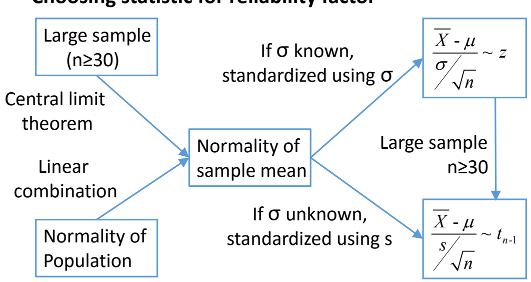
The standard error: $S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{6.6}{\sqrt{41}} = 1.03$

So, the 95% confidence interval for the population mean

= 19 +/- 1.96 x 1.03; or 17.0 to 21.0.

Estimation

Choosing statistic for reliability factor



Estimation

Choosing statistic for reliability factor (Cont.)

When sampling from a:		Reliability factors	
Distribution	Variance	Small sample (n<30)	Large sample (n≥30)
Normal	Known	z-statistic	z-statistic
Normal	Unknown	t-statistic	t-statistic*
Non-normal	Known	Not available	z-statistic
Non-normal	Unknown	Not available	t-statistic*

* *z*-statistic is theoretically acceptable here, but use of the t-statistic is more conservative.

Estimation

Confidence interval estimate (Cont.)

- Confidence interval of population mean with known population variance:

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

- Confidence interval of population mean with unknown population variance:

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

- Degrees of freedom (df) = n-1.

Estimation

Example

- The mean P/E (point estimate) for a sample of 41 firms is 19.0, sample standard deviation is 6.6. What is the 90% confidence interval for the population mean?

Answer:

The standard error: $S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{6.6}{\sqrt{41}} = 1.03$

The t-distribution reliability factor = 1.684 (df=40, α /2=0.05);

So, the 90% confidence interval for the population mean

= 19 +/- 1.684 x 1.03; or 17.27 to 20.73.

Estimation

Confidence interval estimate (Cont.)

- Factors on width of confidence interval:

Factors	Width of confidence interval
Larger confidence level (1-α)	Larger
Larger significance level (α)	Smaller
Larger sample size (n, df)	Smaller
Larger sample standard (s)	Larger
t-distribution (against z-distribution)	Larger

Summary

- **Importance:** ☆☆☆
- **Content:**
  - Central limit theorem and standard error;
  - Confidence interval for population mean and reliability factors.
- **Exam tips:**
  - 常考点1：考概念题，central limit theorem的条件、结论；
  - 常考点2：考计算题，standard error 的计算，构建总体均值置信区间。

Hypothesis Testing (1)

Tasks:

- Describe and interpret the choice of the null and alternative hypotheses;
- Distinguish between one-tailed and two-tailed tests;
- Explain test statistic, significance level, p-value, Type I and Type II errors.

Hypothesis Testing

Estimation Vs. Hypothesis testing

- **Estimation**
  - Addresses the questions such as "what is this parameter's value".
- **Hypothesis testing**
  - **Hypothesis**: a statement about one or more populations;
  - Addresses the questions such as "is the value of the parameter equal to a specific value".

Hypothesis Testing

Steps of hypothesis testing

- Step 1: stating the hypotheses: relation to be tested;
- Step 2: identifying the appropriate test statistic and its probability distribution;
- Step 3: specifying the significance level;
- Step 4: stating the decision rule;
- Step 5: collecting the data and calculating the test statistic;
- Step 6: making the statistical decision;
- Step 7: making the economic or investment decision.

Hypothesis Testing

Null hypothesis vs. Alternative hypothesis

- **Null hypothesis (H_0)**
  - Hypothesis to be tested;
    - ✓ E.g., average monthly return for stock A is equal to 1%.
  - The "=" sign will be only showed in null hypothesis.

Hypothesis Testing

Null hypothesis vs. Alternative hypothesis (Cont.)

- **Alternative hypothesis (H_a)**
  - The opposite side of null hypothesis;
    - ✓ E.g., average monthly return for stock A is not equal to 1%.
  - Hypothesis that analyst wants to approve or conclude;
  - Accepted when the null hypothesis is rejected.

Hypothesis Testing

Test statistic

- A quantity calculated based on a sample.

$$\text{Test statistic} = \frac{\text{Sample statistic-Value of the population parameter under } H_0}{\text{Standard error of the sample statistic}}$$

Critical value (Rejection point)

- A value with which the computed test statistic is compared to decide whether to reject or not reject the null hypothesis.

Hypothesis Testing

Significance level (α)

- The level of significance reflects how much sample evidence we require to reject the null.

p-value

- The smallest level of significance at which the null hypothesis can be rejected.

Hypothesis Testing

Two-tailed test vs. One-tailed test

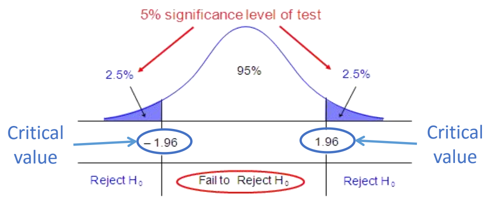
- **Two-tailed test**
  - Used to test if a population parameter is different from a specified value.
    - ✓ $H_0: \theta = \theta_0$ vs. $H_a: \theta \neq \theta_0$
- **One-tailed test**
  - Used to test if a parameter is above or below a specified value.
    - ✓ $H_0: \theta \leq \theta_0$ vs. $H_a: \theta > \theta_0$
    - ✓ $H_0: \theta \geq \theta_0$ vs. $H_a: \theta < \theta_0$

Hypothesis Testing

Two-tailed test vs. One-tailed test (Cont.)

- Example: two-tailed test of population mean

H_0: $\mu = 0$ versus H_a: $\mu \neq 0$

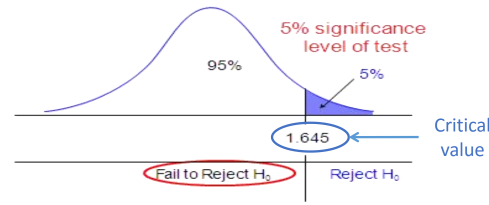


Hypothesis Testing

Two-tailed test vs. One-tailed test (Cont.)

- Example: one-tailed test of population mean

H_0: $\mu \leq 0$ versus H_a: $\mu > 0$



Hypothesis Testing

Decision rule

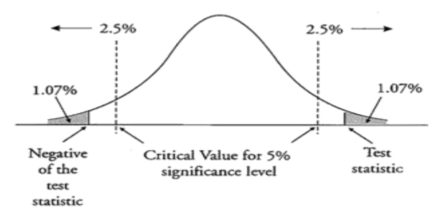
- If test statistic is outside the range of critical value (test statistic ≥ upper critical value, or test statistic ≤ lower critical value), reject the null hypothesis;
- If the p-value is less or equal to the level of significance (α), reject the null hypothesis.

Hypothesis Testing

Decision rule (Cont.)

- An example of decision rule:

Two-tailed hypothesis test with p-value = 2.14%



Hypothesis Testing

Type I error & Type II error

- **Type I error:** rejecting null hypothesis when it is true.
  - P(Type I Error)= Significance level (α).
- **Type II error:** failing to reject the null hypothesis when it is false.
  - P(Type II Error)= β.
- **Power of test:** rejecting the null hypothesis when it is false.
  - Power of test = 1 - P(Type II Error) = 1- β.

Hypothesis Testing

Type I error & Type II error (Cont.)

Decision	True Situation	
	H_0 True	H_0 False
Do not reject H_0	Correct decision	Type II error (Probability = β)
Reject H_0 (accept H_a)	Type I error (Probability = α)	Correct decision (Power of test: 1-β)

Hypothesis Testing

Statistical significance vs. Economic significance

- Statistical significance does not necessarily imply economic significance, due to:
  - Transactions costs;
  - Taxes;
  - Risk.

Summary

- **Importance:** ☆☆☆
- **Content:**
  - Null hypothesis vs. alternative hypothesis;
  - Test statistic & critical value, significance level & p-value, two-tailed test & one-tailed test, type I error and type II error.
- **Exam tips:**
  - 常考点1：null hypothesis和alternative hypothesis的设定；
  - 常考点2：two-tailed test 和 one-tailed test的选择；
  - 常考点3：是否拒绝原假设的decision rules。

Hypothesis Testing (2)

Tasks:

- Identify the appropriate test statistic and interpret the results for a hypothesis test concerning the population mean and variance.

高顿网校 www.gaodun.com

Hypothesis Testing

Hypothesis test concerning a single mean (Cont.)

- Test statistic for hypothesis tests of population mean with Known variance:

$$z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

Where:

- z = z-statistic;
- $\overline{X}$ = sample mean;
- μ_0 = hypothesized value of the population mean;
- σ = population standard deviation.

高顿网校 www.gaodun.com

Hypothesis Testing

Hypothesis test concerning a single mean

- Test statistic for hypothesis tests of population mean with unknown variance:

$$t_{n-1} = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$

Where:

- t_{n-1} = t-statistic with n-1 degrees of freedom (n is the sample size);
- $\overline{X}$ = sample mean;
- μ_0 = hypothesized value of the population mean;
- s = sample standard deviation.

高顿网校 www.gaodun.com

Hypothesis Testing

Hypothesis test concerning a single mean (Cont.)

- Test statistic for hypothesis tests of population mean with unknown variance and large sample size:

$$z = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$

Where:

- z = z-statistic;
- $\overline{X}$ = sample mean;
- μ_0 = hypothesized value of the population mean;
- s = sample standard deviation.

Hypothesis Testing

Example (1)

- Test the hypothesis that a fund's mean return is equal to 1% per month at 5% significance level, the population is normal distributed. The data provided is:
  - Sample mean: 1.5%;
  - Sample size: 45;
  - Standard deviation of population: 1.4%.

Hypothesis Testing

Answer (1)

- Step 1: H_0: $\mu = 0.01$ and H_a: $\mu \neq 0.01$;
- Step 2: with known population variance (standard deviation), use two-tailed *z*-test;
- Step 3: The critical *z*-values for 5% significance level (95% confidence interval) are +/- 1.96;
- Step 4: decision rule: if the z-statistic is outside the range of critical values (-1.96 to +1.96), reject H_0;

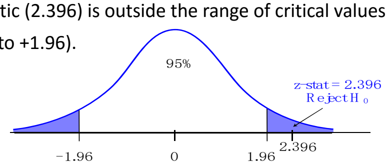
Hypothesis Testing

Answer (1)

- Step 5: calculate the test statistic;

$$\text{z-statistic} = \frac{0.015-0.01}{0.014/\sqrt{45}} = 2.396$$

- Step 6: reject H_0 (mean return = 1%), because the *z*-statistic (2.396) is outside the range of critical values (-1.96 to +1.96).



Hypothesis Testing

Example (2)

- Researcher believes a fund's mean returns (μ_{Fund}) **exceed** 1% per month. Sample size is 36, sample mean is 1.5%, and sample standard deviation is 1.8%. The population is normal distributed. Test the null hypothesis at 5% significance level.

Hypothesis Testing

Answer (2)

- Step 1: H_0: $\mu \le 0.01$ and H_a: $\mu > 0.01$;
- Step 2: with unknown population variance and large sample size (36), use one-tailed *z*-test (right tail);
- Step 3: The critical *z*-value for 5% significance level is 1.65;
- Step 4: decision rule: if the z-statistic is above the critical values (1.65), reject H_0;
- Step 5: calculate the test statistic: $z=\dfrac{0.015-0.01}{0.018/\sqrt{36}}=1.667$
- Step 6: reject the null hypothesis.

Hypothesis Testing

Hypothesis test concerning difference of means

- Null hypotheses and alternative hypothesis:
  - H_0: $\mu_1 - \mu_2 = d_0$ and H_a: $\mu_1 - \mu_2 \neq d_0$;
  - H_0: $\mu_1 - \mu_2 \ge d_0$ and H_a: $\mu_1 - \mu_2 < d_0$;
  - H_0: $\mu_1 - \mu_2 \le d_0$ and H_a: $\mu_1 - \mu_2 > d_0$.

Hypothesis Testing

Hypothesis test concerning difference of means (Cont.)

- Test statistic for hypothesis tests of the difference of two independent population means with variance unknown but assumed equal.

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\left(\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}\right)^{1/2}} \quad \text{where: } s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$df = n_1 + n_2 - 2$$

Hypothesis Testing

Hypothesis test concerning difference of means (Cont.)

- Test statistic for hypothesis tests of the difference of two independent population means with variance unknown but assumed unequal.

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^{1/2}} \quad \text{where: } df=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1}+\frac{(s_2^2/n_2)^2}{n_2}}$$

Hypothesis Testing

Hypothesis test concerning mean differences

- Also referred to as paired comparison test;
- Null hypotheses and alternative hypothesis:
  - H_0: $\mu_d = \mu_0$ and H_a: $\mu_d \neq \mu_0$;
  - H_0: $\mu_d \geq \mu_0$ and H_a: $\mu_d < \mu_0$;
  - H_0: $\mu_d \leq \mu_0$ and H_a: $\mu_d > \mu_0$.

Hypothesis Testing

Hypothesis test concerning mean differences (Cont.)

- Test statistic for hypothesis tests of the mean differences between two dependent populations with unknown variance.

$$t = \frac{\bar{d} - \mu_{d0}}{S_{\bar{d}}}$$

where: n = the number of paired observations;
$\bar{d}$ = sample mean difference;
$s_{\bar{d}}$ = the standard error of $\bar{d}$;
df = n-1.

Hypothesis Testing

Hypothesis test concerning a single variance

- Null hypotheses and alternative hypothesis:
  - H_0: $\sigma = \sigma_0$ and H_a: $\sigma \neq \sigma_0$;
  - H_0: $\sigma \geq \sigma_0$ and H_a: $\sigma < \sigma_0$;
  - H_0: $\sigma \leq \sigma_0$ and H_a: $\sigma > \sigma_0$.

Hypothesis Testing

Hypothesis test concerning a single variance

- Test statistic for hypothesis tests of a value of a population variance.

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2} \quad \text{(Chi-Square)}$$

where: n = sample size;
s^2 = sample variance;
σ_0^2 = the hypothesized value;
df = n-1.

Hypothesis Testing

Hypothesis test concerning equality of two variances

- Null hypotheses and alternative hypothesis:
  - H_0: $\sigma_1 = \sigma_2$ and H_a: $\sigma_1 \neq \sigma_2$;
- Test statistic for hypothesis tests of equality of two population variances:

$$F = \frac{S_1^2}{S_2^2} \text{ with df of } (n_1-1,\ n_2-1)$$

where: n_1 = the number of large sample size;
n_2 = the number of small sample size;
s_1^2 = the large sample variance in numerator;
s_2^2 = the small sample variance in denominator.

Hypothesis Testing

Parametric tests vs. Nonparametric tests

- Parametric tests
  - Based on assumptions about population distributions and population parameters.
    - ✓ E.g., t-test, z-test, F-test.

Hypothesis Testing

Parametric tests vs. Nonparametric tests (Cont.)

- Nonparametric tests
  - Test things other than parameter values.
  - Applied when:
    - ✓ Data do not meet distributional assumptions;
    - ✓ Data are given in ranks;
    - ✓ The hypothesis we are addressing does not concern a parameter.

Summary

- **Importance:** ☆☆☆
- **Content:**
  - Test statistic for population mean with known variance (z) and unknown variance (t), difference of means (t), mean differences (t), and population variance (χ^2), equality of two variance (F).
- **Exam tips:**
  - 常考点：根据检验的内容选择假设检验的种类。

Technical Analysis

Tasks:

- Explain principles of technical analysis, its applications, and its underlying assumptions;
- Describe common chart patterns and indicators;
- Describe the usage of cycles, Elliott wave theory, and Fibonacci numbers.

Overview of Technical Analysis

Principles of technical analysis

- Analyzed using price and volume;
- Prices are determined by supply and demand;
- Market reflects the collective knowledge and sentiment of many varied participants and the amount of buying and selling activity in a particular security.

Overview of Technical Analysis

Assumptions of technical analysis

- The trades determine volume and price;
- Market price reflects both rational and irrational behavior of market participants;
- The Efficient Market Hypothesis (EMH) does not hold;
- The securities are freely traded in the market;
- The trends and patterns tend to repeat themselves which makes the price predictable.

Overview of Technical Analysis

Advantages of technical analysis

- It is based on actual trade data;
- It can be used for assets with no cash flows to be discounted for valuation (e.g., commodities, currencies).

Disadvantages of technical analysis

- May not work in illiquid markets;
- May not work in markets subject to manipulation;
- May not work for valuing bankrupt companies.
  - Stock price → 0, but short covering may create positive technical patterns.

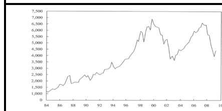
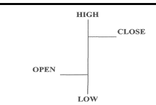
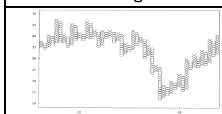
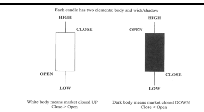
Overview of Technical Analysis

Technical analysis vs. Fundamental analysis

Technical analysis	Fundamental analysis
Studying the price and volume data	Numerous estimates and assumptions are added in financial statements
Trying to find the price it will trade	Trying to find the intrinsic value it should trade
It works in commodities markets	It may not work in commodities markets

Technical Analysis Charts

Type of charts

Line chart	Bar chart
	
Point and figure chart	**Candlestick chart**
	

Technical Analysis Charts

Patterns of charts

- **Trend**
  - Uptrend: the price goes to higher highs and higher lows;
  - Downtrend: the price goes to the lower lows and lower highs.
- **Support line**
  - A low price range in which buying activity is sufficient to stop the decline in price.

Technical Analysis Charts

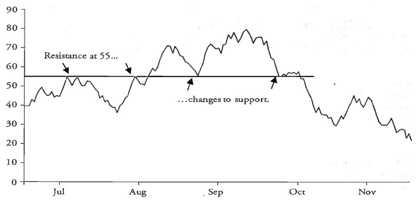
Patterns of charts (Cont.)

- **Resistance line**
  - A price range in which selling activity is sufficient to stop the rise in price.

Technical Analysis Charts

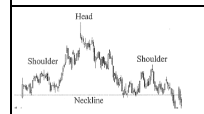
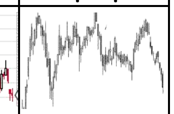
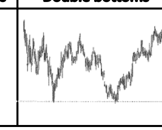
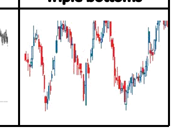
Patterns of charts (Cont.)

- **Change in polarity**
  - Once support level is breached, it becomes resistance level;
  - Once resistance level is breached, it becomes support level.



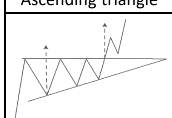
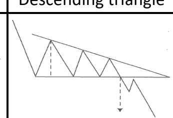
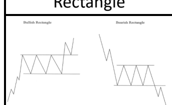
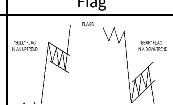
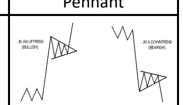
Technical Analysis Charts

Patterns of charts - reverse patterns

Head and shoulders	Double tops	Triple tops
		
Inverse head and shoulders	**Double bottoms**	**Triple bottoms**
		

Technical Analysis Charts

Patterns of charts - continuing patterns

Ascending triangle	Descending triangle	Symmetrical triangle
		
Rectangle	Flag	Pennant
		

Technical Analysis Indicators

Type of technical analysis indicators

Price-based Indicators	Moving average and Bollinger Bands
Momentum Indicators	ROC oscillator, RSI, Stochastic Oscillator, and MACD
Sentiment Indicators	Opinion Polls, Put-call ratio, VIX, and short interest ratio
Flow of Funds Indicators	Arms Index, Margin Debt, Mutual Fund Cash Position, New Equity Issuance, and Secondary Offerings

Cycle Analysis

Cycle analysis

- Kondratieff Wave: 54-year cycle;
- 18-year cycle;
- Decennial cycle: 10 years;
- Presidential cycle (U.S.): 4 years.
  - 1st and 2nd year: worst performance;
  - 3rd and 4th year: best performance.

Elliott Wave Theory

Elliott wave theory

- Market prices move in interconnected cycles that range from very short-term to very long-term;
  - Uptrends: 5 waves up, 3 waves down;
  - Downtrends: 5 waves down, 3 waves up.
- Wave sizes conform to **Fibonacci sequence.**
  - 0, 1, 1, 2, 3, 5, 8, 13, 21 , 34, ……
  - Add last two numbers to get next number;
  - The ratio of one Fibonacci number to the next number converge around 0.618.

Inter-market Analysis

Inter-market analysis

- Inter-market analysis is based on the principle that all markets are interrelated and influence each other.
  - Involves the use of relative strength analysis for different groups of securities to make allocation decisions.
    - ✓ E.g., stocks versus bonds, sectors in an economy, and securities from different countries.

Summary

- **Importance:** ☆
- **Content:**
  - Principles, consumptions, advantage & disadvantage of technical analysis;
  - Price charts, indicators, cycles, and wave theory.
- **Exam tips:**
  - 不是重要考点。