Terms

- **Random experiment (随机试验)**
  - An observation or measurement process with multiple but uncertain outcomes.
- **Outcome (结果)**
  - The result of a single trial. For example, if we roll two dices, an outcome might be 3 and 4; a different outcome might be 5 and 2.

- **Random variable:** a variable whose possible values are outcomes of a random phenomenon.
  - Discrete random variables
  - Continuous random variables
- **Event (事件)**
  - ✓ Mutually exclusive events (互斥事件): Events that cannot both happen at the same time.
  - ✓ Exhaustive events (完备事件): Those include all possible outcomes.

Terms

1 ➢ **Random experiment (随机试验)**

- An observation or measurement process with multiple but uncertain outcomes.

3 ➢ **Outcome (结果)**

- The result of a single trial. For example, if we roll two dices, an outcome might be 3 and 4; a different outcome might be 5 and 2.

2 ➢ **Random variable:** a variable whose possible values are outcomes of a random phenomenon. 取值不确定的量

- Discrete random variables
- Continuous random variables

➢ **Event (事件)**

✓ Mutually exclusive events (互斥事件): Events that cannot both happen at the same time.

✓ Exhaustive events (完备事件): Those include all possible outcomes.

Terms

1 ➢ **Random experiment (随机试验)**

- An observation or measurement process with multiple but uncertain outcomes.

3 ➢ **Outcome (结果)**：一个确定的数

- The result of a single trial. For example, if we roll two dices, an outcome might be 3 and 4; a different outcome might be 5 and 2.

2 ➢ **Random variable:** a variable whose possible values are outcomes of a random phenomenon. 取值不确定的量 X

- Discrete random variables
- Continuous random variables

4 ➢ **Event (事件)** = 2+3 X=1

✓ Mutually exclusive events (互斥事件): Events that cannot both happen at the same time.

✓ Exhaustive events (完备事件): Those include all possible outcomes.

遍历事件

Probability

一、

1、定义：

➢ **Probability of an event**

$P(X=1) = \frac{1}{6}$

$$P(A) = \frac{number\ of\ outcomes\ favorable\ to\ A}{total\ number\ of\ outcomes}$$

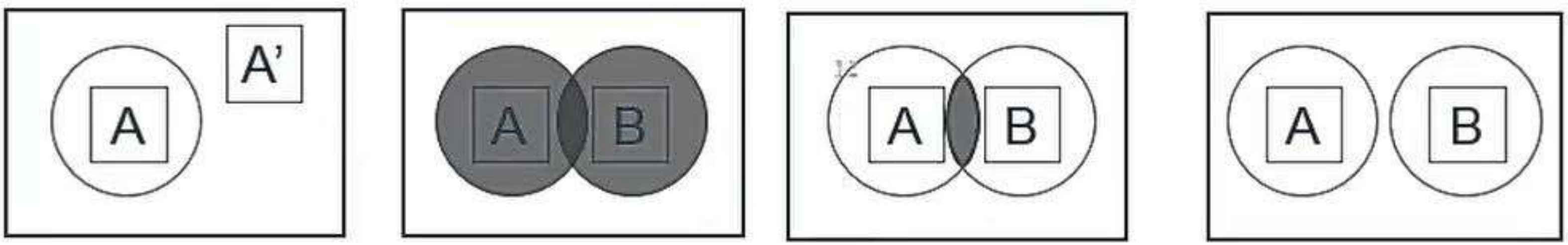
2.性质

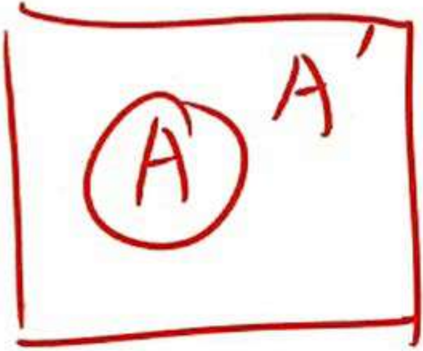
➢ **Two defining properties of probability**

- $0 \leq P(E) \leq 1$
- If E_1, E_2, ……, E_n is mutually exclusive and exhaustive, then:

  $P(E_1) + P(E_2) + …… + P(E_n) = 1$

- Venn Diagrams





A'：不是A.

A 与 A' 互斥

Probability

3. 分类

- **Joint Probability**
  - The probability that the random variables (in this case, both random variables) take on certain values simultaneously, P(AB).
- **Unconditional Probability (边际概率, a.k.a marginal probability)**
  - The expected value of the variable without any restrictions (or lacking any prior information), P(A).
- **Conditional Probability (条件概率)**
  - An expected value for the variable conditional on prior information or some restriction (e.g., the value of a correlated variable). The conditional expectation of B, conditional on A , is given by P(B|A).

Probability

3. 分类

- **Joint Probability** 联合概率 交集：A、B同时发生
  - The probability that the random variables (in this case, both random variables) take on certain values simultaneously, P(AB).
- 非条件概率 **Unconditional Probability (边际概率, a.k.a marginal probability)**
  - The expected value of the variable without any restrictions (or lacking any prior information), P(A). = 50%
- **Conditional Probability (条件概率)** 在经济状况好的条件下 stock price个
  - An expected value for the variable conditional on prior information or some restriction (e.g., the value of a correlated variable). The conditional expectation of B, conditional on A , is given by P(B|A) 事件 >条件：在A发生的条件下，B发生的概率

$P(A \underline{\text{and}} B) = \underline{P(AB)}$

和

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

或

Probability and Probability Distributions

- **Unconditional probability: P(A), P(B)** 关系
- **Conditional probability: P(A|B)**

$$P(A|B) = \frac{P(AB)}{P(B)}; P(B) > 0$$

$$P(B|A) = \frac{P(AB)}{P(A)}; P(A) > 0$$

飞

北京 火 上海 飞 火 船 深圳

2×3.

乘法法则：P(AB) = P(A)·P(B|A)

= P(B)·P(A|B)

Probability and Probability Distributions

独立事件：定义

- **The occurrence of A has no influence 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 strongly.

Probability and Probability Distributions

独立事件：定义

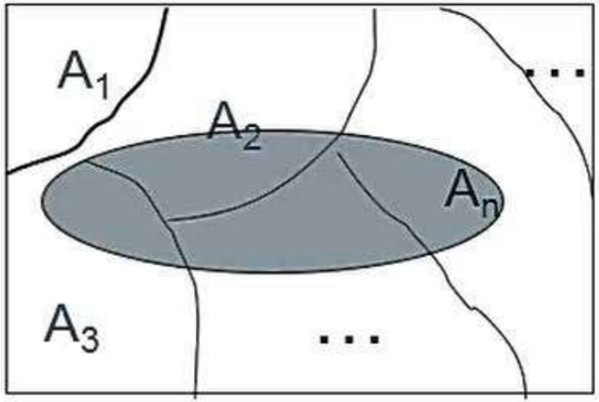
- **The occurrence of A has no influence of on the occurrence of B.**
  - $P(A|B) = P(A)$ or $P(B|A) = P(B)$ √
  - ※ $P(AB) = P(A) \times P(B)$ 独立. √
  - $P(A \text{ or } B) = P(A) + P(B) - P(AB)$ 加法法则
- **Independence and Mutually Exclusive are quite different.**
  - If exclusive, must not independence
    - ✓ Cause exclusive means if A occur, B can not occur, A influents B strongly.

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

四 Probability and Probability Distributions

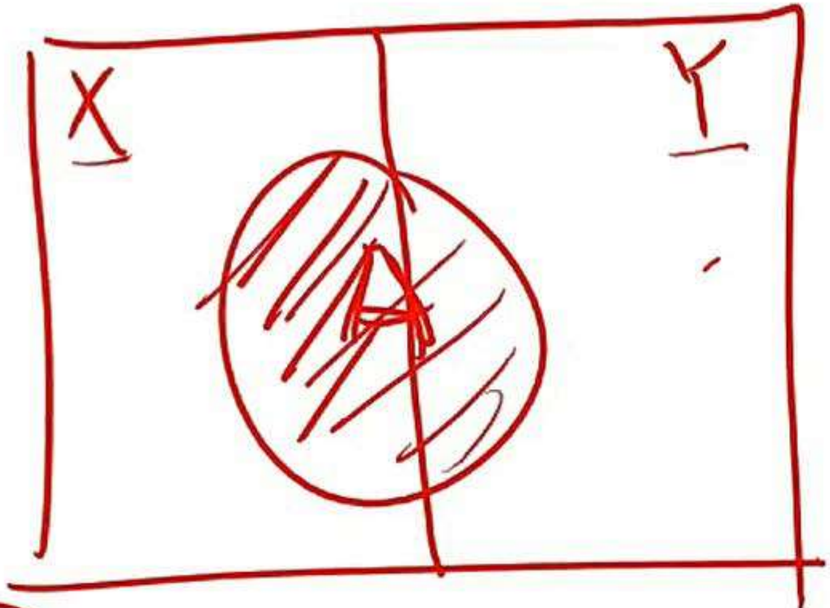
- **Total Probability Formula** 全概率公式：①用途：用来计算非条件概率

  - If an event A must result in one of the mutually exclusive events $A_1, A_2, A_3, \ldots\ldots, A_n$, then
  - $P(A) = P(A_1)P(A|A_1) + P(A_2)P(A|A_2) + \ldots + P(A_n)P(A|A_n)$



(1) $A_i A_j = \Phi \ (i \neq j)$

(2) $\bigcup_{i=1}^{n} A_i = \Omega$



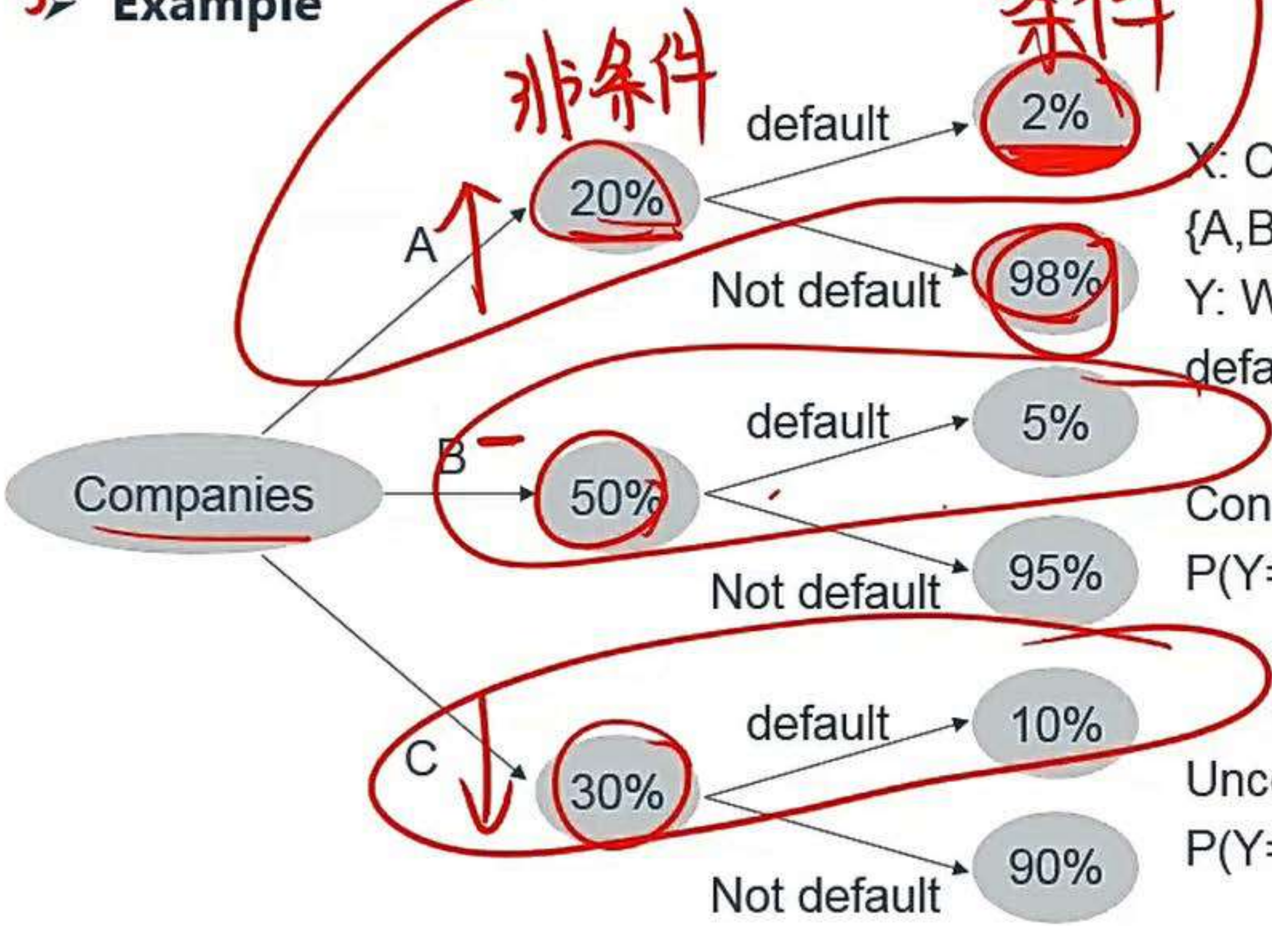
$$P(A) = P(AX) + P(AY) \checkmark$$

$$= P(X)\cdot P(A|X) + P(Y)\cdot P(A|Y)$$

全概率公式

Probability and Probability Distributions

Example



X: Company's choice – {A,B,C}

Y: Whether company will default – {0,1}

Conditional Probability:

P(Y=1|X=A)=?

Unconditional Probability:

P(Y=1)=?

P(D) =

Probability and Probability Distributions

2. Bayes' Theorem 贝叶斯公式：用来计算条件概率

$$P(A|B) = \frac{P(B|A)}{P(B)} \times P(A) \longleftarrow \text{Prior Probability}$$

➤ **Example**

- 一个人有病的概率是10%，没病的概率是90%。在有病的情况下机器诊断出有病的概率是99%，诊断出没病的概率是1%；在没病的情况下机器诊断出有病的概率是5%，诊断出没病的概率是95%。若机器诊断出有病的情况下人真的有病的概率是多少？

	机器说有病	机器说没病
如果人真有病	0.99	0.01
如果人真没病	0.05	0.95

Probability and Probability Distributions

2. Bayes' Theorem 贝叶斯公式：用来计算条件概率

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

99%
10人

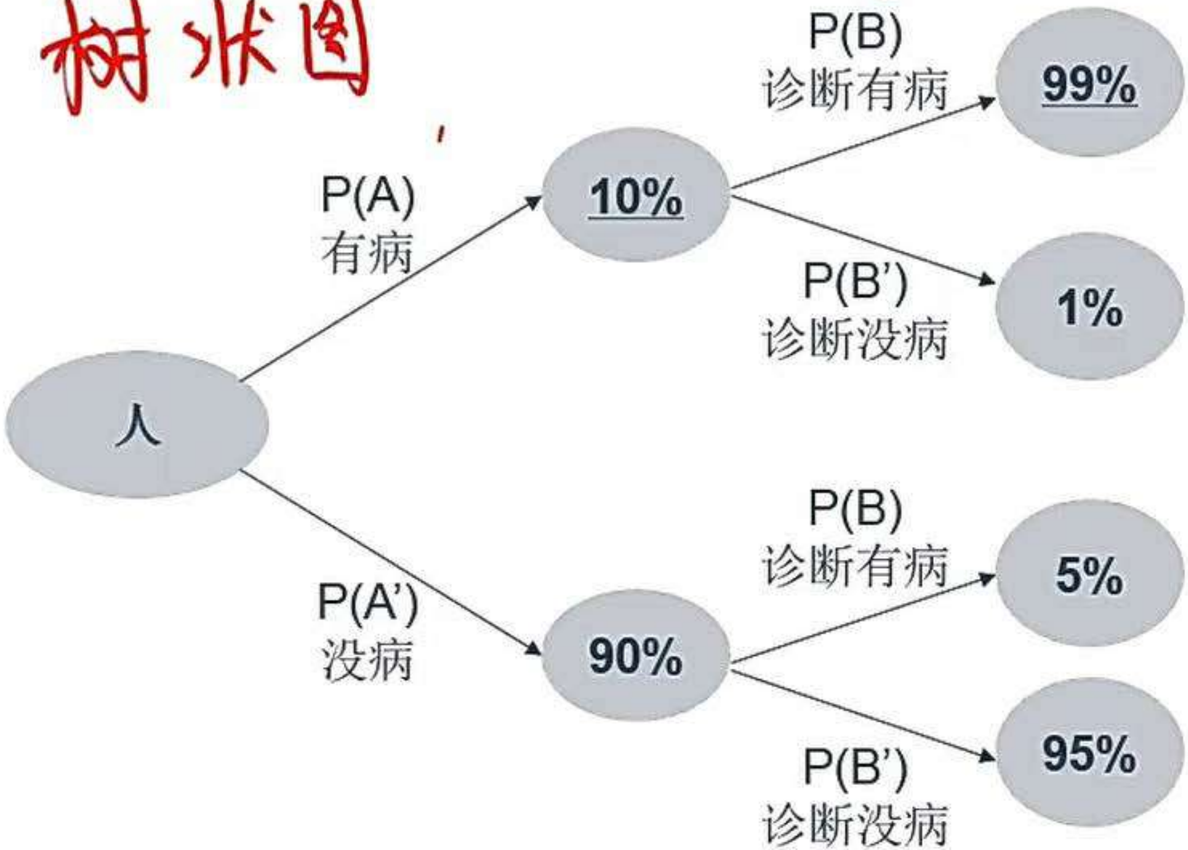
➢ Example $\cancel{P(AB)} = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$

- 一个人有病的概率是10%，没病的概率是90%。在有病的情况下机器诊断出有病的概率是99%，诊断出没病的概率是1%；在没病的情况下机器诊断出有病的概率是5%，诊断出没病的概率是95%。若机器诊断出有病的情况下人真的有病的概率是多少？

	机器说有病	机器说没病
如果人真有病	0.99	0.01
如果人真没病	0.05	0.95

Probability and Probability Distributions

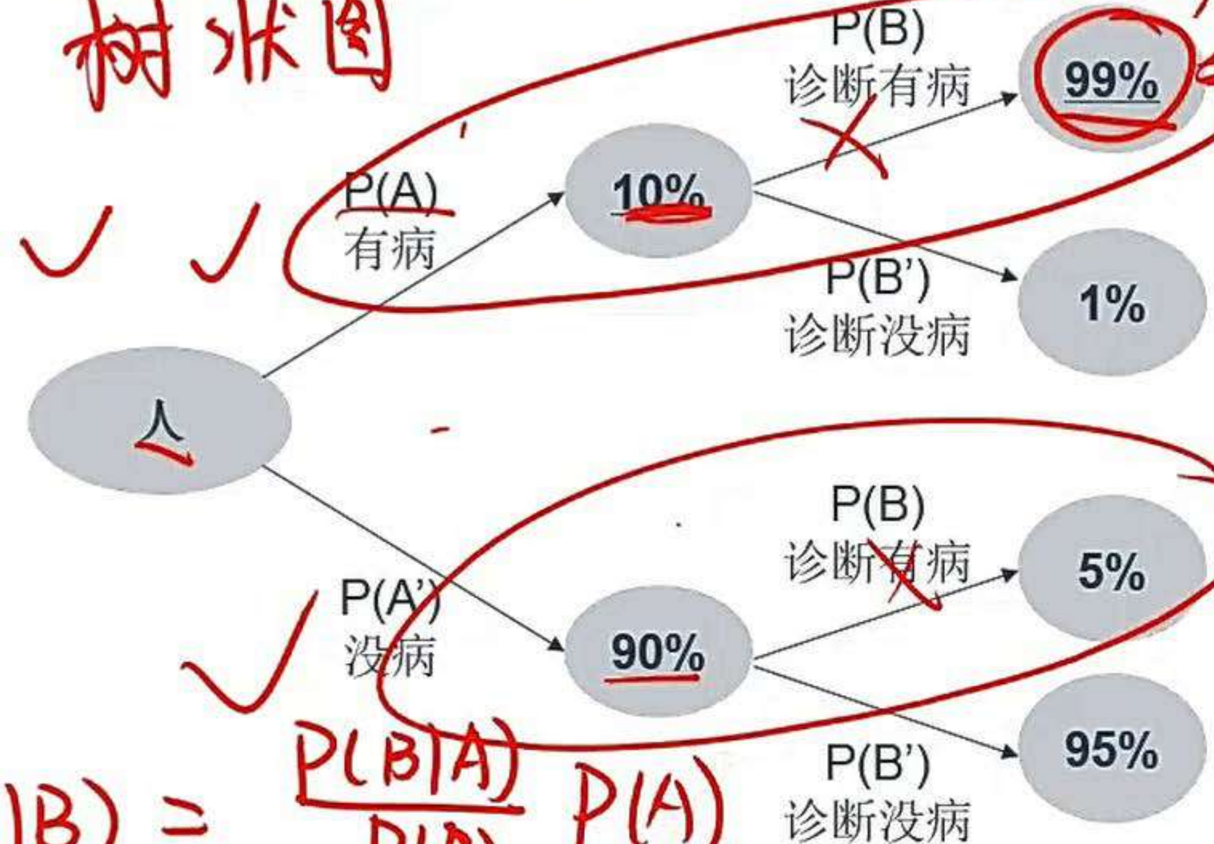
树状图



$$P(A|B) = \frac{99\% \times 10\%}{99\% \times 10\% + 5\% \times 90\%} = 68.75\%$$

Probability and Probability Distributions

树状图



条件 $P(B|A)$

$P(\bar{B}|A)$

$P(B|\bar{A})$

$P(\bar{B}|\bar{A})$

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

$$P(A|B) = \frac{99\% \times 10\%}{99\% \times 10\% + 5\% \times 90\%} = 68.75\%$$

三、形式：图、表、函数

图.

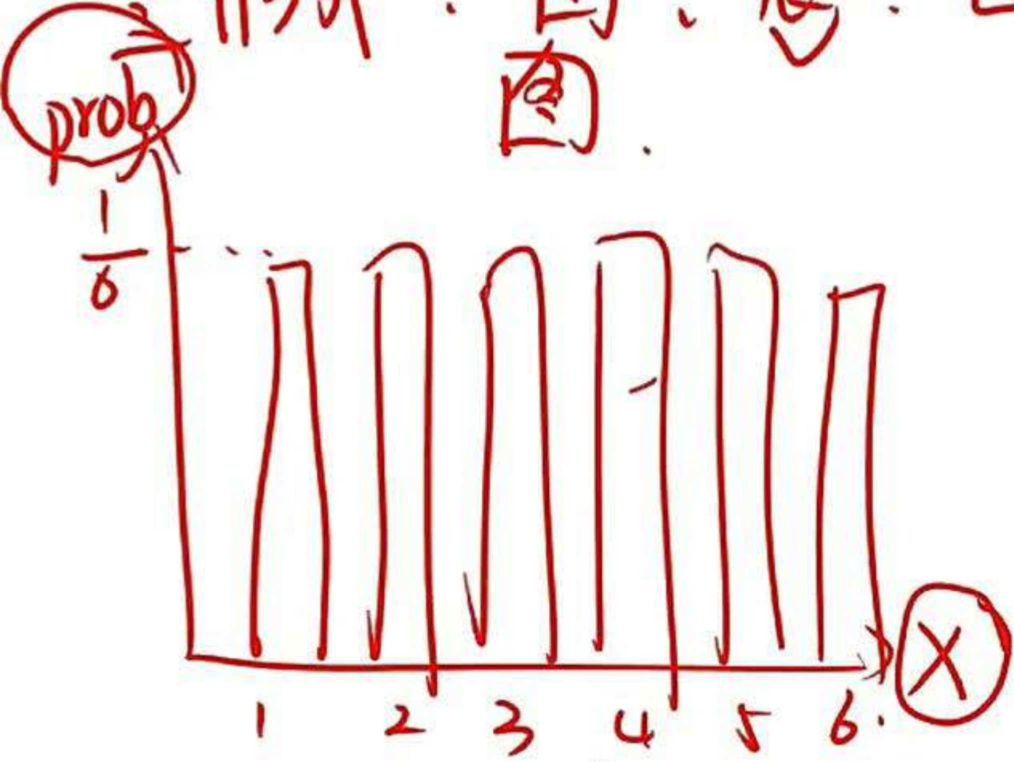


表.

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$f(X = x_i) = \frac{1}{6} \quad (x_i = 1, 2, 3, 4, 5, 6)$$

Probability and Probability Distributions

四 prob. function.

- **Probability Distribution** ：三种形式. $0 \le P \le 1$
  - Describe the probabilities of all the possible outcomes for a random variable.

离散 连续的

- **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.
    - ✓ $P(x_1 < X < x_2)$

X_1 X_2

$\frac{X_1+X_2}{2}$

Probability and Probability Distributions

➢ Random Variables and Their Probability Distributions

- Probability Distribution of a Discrete Random Variable
  - ✓ Probability Mass Function (PMF) or Probability Function (PF)
  - ✓ $f(X = x_i) = P(X = x_i),\ i = 1,2,3\ldots$

$P(x=7)=0$

性质

- ✓ Properties of the PMF

$$1.\ f(X = x_i) = 0,\ x \neq x_i$$

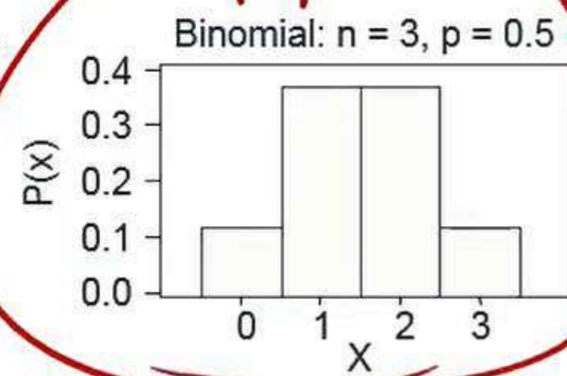
$$2.\ 0 \leq f(x_i) \leq 1$$

$$3.\ \sum_{x} f(x_i) = 1$$

For example:
Binomial
n = 3 p = 0.5

x	P(x)
0	0.125
1	0.375
2	0.375
3	0.125
	1.000

PF



$P(x_1 \le X < x_2)$ 与 $P(x_1 \le X \le x_2)$.

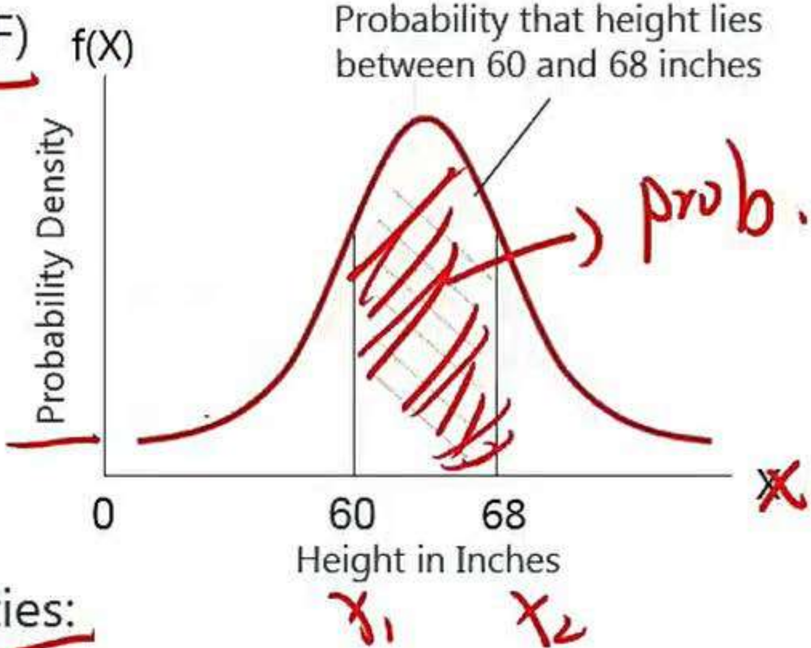
X continuing $=$

Probability and Probability Distributions

➢ Probability Distribution of a Continuous Random Variable

- Probability density function (PDF)

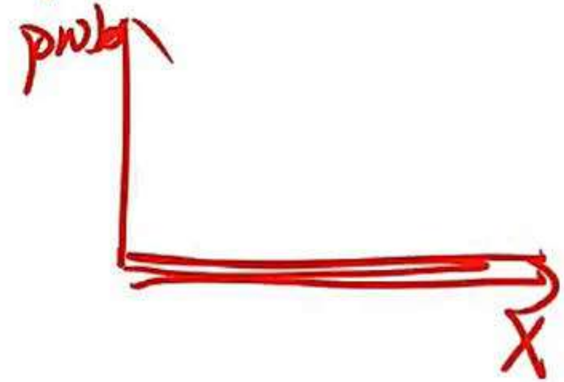
$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$$



- A PDF has the following properties:
  - ✓ The total area under the curve f(x) is 1. ✓
  - ✓ $P(x_1 < X < x_2)$ is the area under the curve between x_1 and x_2.
  - ✓ $P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2)$

Probability and Probability Distributions

- **Probability function: p(x) = P(X = x)** PF / PMF mass
  研究对象
  - 1、For discrete random variables
  - $0 \le p(x) \le 1$
  - $\Sigma p(x) = 1$
- 2、**Probability density function (p.d.f): f(x)** 概率密度函数.
  ①研究对象
  - For continuous random variable commonly

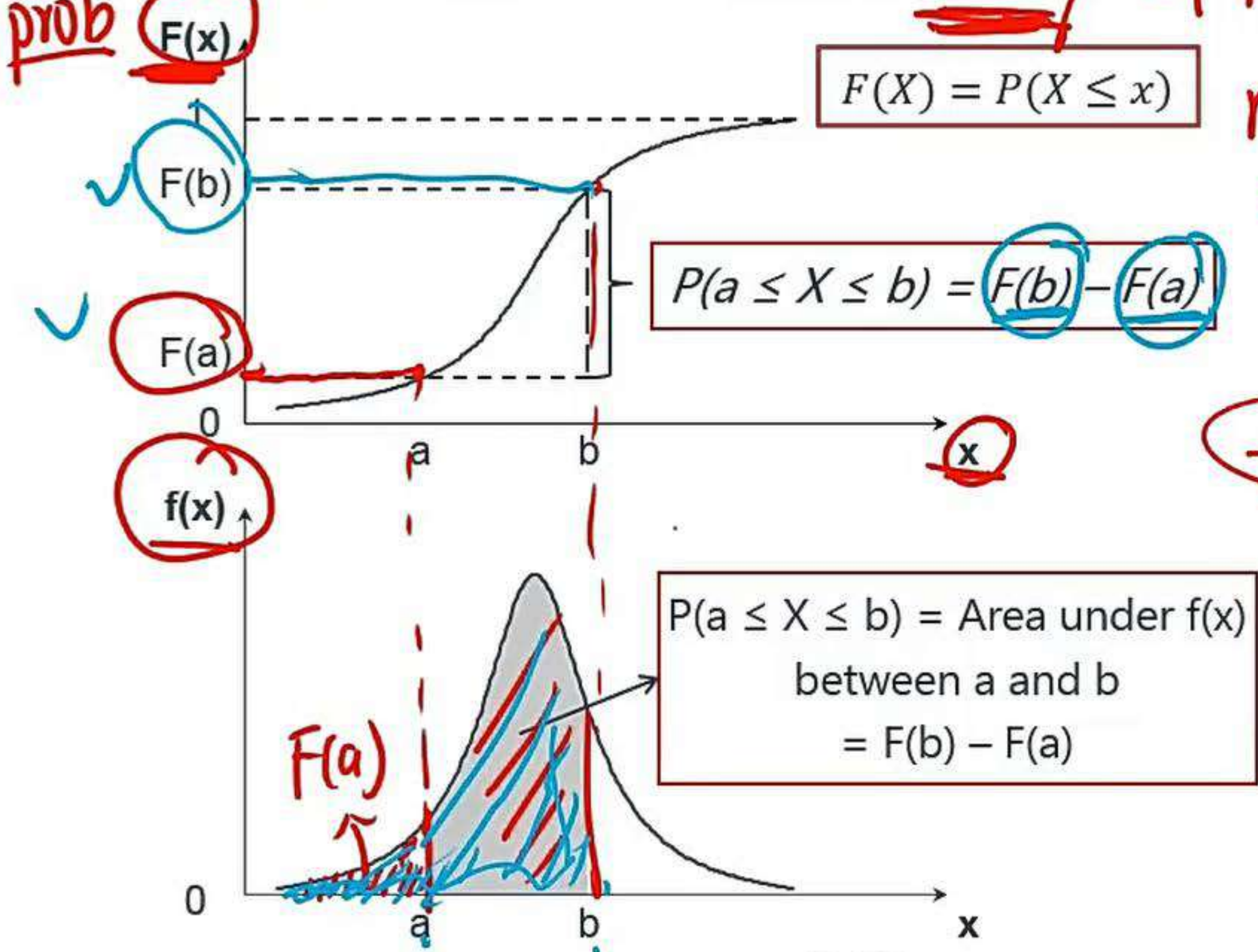


- **Cumulative probability function (c.p.f): F(x)**
  - Discrete
    - ✓ $\boxed{F(x) = P(X \le x)}$
  - Continuous
    - ✓ $\boxed{F(x) = \int_{-\infty}^{x} (u)du}$

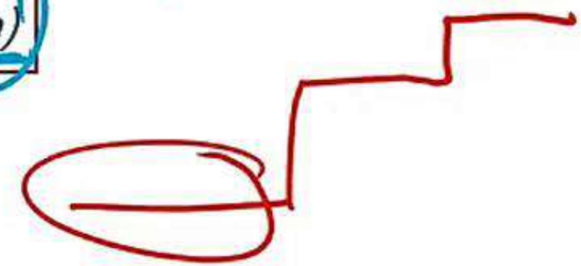
Probability and Probability Distributions

- **Cumulative Distribution Function (CDF)**

CDF ①研究对象



non-decreasing
非递减函数



Probability and Probability Distributions

- **Properties of CDF**
  - $F(-\infty) = 0$ and $F(+\infty) = 1$
  - $F(X)$ is a non-decreasing function such that if $x_2 > x_1$ then $F(x_2) \geq F(x_1)$.
  - $P(X \geq k) = 1 - F(k)$
  - $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$

Probability and Probability Distributions

- **Properties of CDF**
  - $F(-\infty) = 0$ and $F(+\infty) = 1$
  - CDF
  - $F(X)$ is a non-decreasing function such that if $x_2 > x_1$ then $F(x_2) \geq F(x_1)$.
  - $P(X \geq k) = 1 - F(k)$ ∨ CDF：从左到右的面积
  - $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$

Probability and Probability Distributions

- **Multivariate probability density function**

eg:
  - We take X from 1 or 2 with the same probability. We take Y from [1,X] with the same probability.

X = 1 50%
2 50%

Y	X		
	1	2	Total
1	0.50	0.25	0.75
2	0.00	0.25	0.25
Total	0.50	0.50	1.00

联合概率

非条件

- **Definition: f(X,Y) = P(X=x and Y=y)**
- **Properties of the bivariate or joint probability mass function (PMF)**
  - $f(X,Y) \geq 0$ for all pairs of X and Y. This is because all probabilities are nonnegative.
  - $\sum\sum f(X,Y) = 1$.

Probability and Probability Distributions

➢ **Marginal probability function**

Marginal probability distribution of X and Y			
Value of X	f(x)	Value of Y	f(y)
1	0.50	1	0.75
2	0.50	2	0.25
Sum	1.00		1.00

➢ **Definition of marginal probability function**

$$f(X) = \sum_{y} f(X,Y) \text{ for all X}$$

$$f(Y) = \sum_{x} f(X,Y) \text{ for all Y}$$

21-225

Probability and Probability Distributions

- **Statistical Independence** 如何证明统计上的独立
  - Definition of Statistical Independence: f(X,Y) = f(X)f(Y)

Y \ X	1	2	3	f(Y)
1	1/9	1/9	1/9	3/9
2	1/9	1/9	1/9	3/9
3	1/9	1/9	1/9	3/9
f(X)	3/9	3/9	3/9	1

= P(Y=1) 边缘条件

P(X=1)

$$P(AB) = P(A) \cdot P(B).$$

$$P(X=1, Y=1) = P(X=1) \cdot P(Y=1)$$

$$\frac{1}{9} = \frac{3}{9} \times \frac{3}{9}$$

P(X

Example 1

X

B

- The joint probability distribution of random variables X and Y is given by $f(x,y) = kxy$ for $x = 1,2,3$, $y = 1,2,3$ and k is a positive constant, what is the probability that X+Y will exceed 5?

常数

A. 1/9

B. 1/4

C. 1/36

D. Cannot be determined.

$X+Y>5$

$X=3 \quad Y=3$.

$f(3,3) = k\times3\times3$

- Correct Answer : B

Y \ X	1	2	3	
1	k	2k	3k	6k
2	2k	4k	6k	12k
3	3k	6k	9k	18k

$36k=1 \quad k=\frac{1}{36}$

总结：

1、三个事件
- 互斥：二者不能同时发生
- 完备：包含所有可能性
- 独立：A的发生不受B的影响

2 三个prob.
- 非条件：不考虑其他因素 $P(A)$
- 条件：考虑其他因素 $P(B|A)$ 事件 条件
- 联合：A.B同时发生 $P(AB)$

三者之间的关系：$P(AB)=P(A)\cdot P(B|A)=P(B)\cdot P(A|B)$

3. 两个 formula

- 全概率
  - 1. 定义：计算非条件概率
  - 2. 公式：(推导)
  - 3. 目的：为贝叶斯服务
- 贝叶斯
  - 1. 定义：计算条件概率
  - 2. 公式：(乘法法则)
  - 3. 计算：两个方式
    - 图
    - 公式

4 三个 prob. function

- PF / PMF
  - 1. 研究对象
  - 2. 图象
  - 3. 性质
- PDF
  - 1. 研究对象
  - 2. prob. density
- CDF
  - 1. 研究对象
  - 2. 非递减

prob. density

3. 两个 formula {全概率 {1.定义：计算概率
2.公式：(摘要)
3.目的：为贝叶斯服务
贝叶斯 {1.定义：计算条件概率
2.公式：(乘法法则)
3.计算：两个方式 {图
公式

$F(X \geq k) = 1 - F(k)$

4 三个 prob. function {PF/PMF {1.研究对象
2.图象
3.性质
PDF {1.研究对象
2.prob. density
3.
CDF {1.研究对象
2.非递减
3.

prob. density

5. 二维PDF eg：例题。

6. 证明随机变量之间是相互独立。

$$P(AB) = P(A) \cdot P(B).$$

Basic Statistics 基本统计量

矩：

一阶矩：mean

二阶矩：variance. covariance correlation

三阶矩：skewness 偏度

四阶矩：

$$\boxed{E(\underline{x} - \underline{\mu}) = E(\underline{x}) - \underline{\mu} = 0}$$

Central Moment $E(X-\mu)$

- **Moment** 矩
  - The k-th moment of X is defined as: $m_K = E(X^K)$ $E(X^1)$ μ
  - If k = 1, then $m_1 = E[X]$, it is the mean.
- **Central moment** 中心矩
  - The k-th central moment of X is defined as: $\mu_K = E[(X-\mu)^K]$
  - Central moments are measured relative to the mean.
  - If k = 1, the first central moment is equal to 0.
  - If k = 2, the second central moment is the variance $= E(X-\mu)^2$
  - If k = 3, then the third central moment divided by the cube of the standard deviation is the skewness. $= \frac{E(X-\mu)^3}{\sigma^3}$
  - If k = 4, then the fourth central moment divided by the square of the variance is the kurtosis. $= \frac{E(X-\mu)^4}{\sigma^4}$

12

Expected Value 一阶矩

- **Expected Value** ：1.定义：衡量数据的中心趋势
  - A measure of central tendency – the first moment.

2公式：

$$E(X) = \sum P(x_i)\,x_i = P(x_1)x_1 + P(x_2)x_2 + \ldots + P(x_n)x_n$$ 加权平均.

$$\frac{x_1\ x_2\ x_3}{}$$

$$E(X) = \int xf(x)dx$$

$$\sqrt[3]{x_1 x_2 x_3}$$

- **Properties of Expected Value**
  - If b is a constant, E(b) = b.
  - If a is a constant, E(aX) = aE(X).
  - If a and b are constants, then E(aX + b) = aE(X) + E(b) = aE(X) + b.
  - $E(X^2) \neq [E(X)]^2$
  - E(X + Y) = E(X) + E(Y)
  - In general, E(XY) ≠ E(X)E(Y); If X and Y are independent random variables, then E(XY) = E(X)E(Y).

一 Expected Value 一阶矩

- **Expected Value** ：1.定义：衡量数据的中心趋势
  - A measure of central tendency – the first moment.

2.公式：

$$E(X) = \sum P(x_i)\, x_i = P(x_1)x_1 + P(x_2)x_2 + \ldots + P(x_n)x_n$$ 加权平均.

$$E(X) = \int x f(x)\, dx$$

$\frac{x_1\ x_2\ x_3}{3}$

$\sqrt[3]{x_1 x_2 x_3}$

3.性质

- **Properties of Expected Value**
  - If b is a constant, E(b) = b.
  - If a is a constant, E(aX) = aE(X).
  - If a and b are constants, then E(aX + b) = aE(X) + E(b) = aE(X) + b.
  - ☆ $\boxed{E(X^2) \neq [E(X)]^2}$ = variace $E(X-\mu) = E(X) - \mu = 0$
  - E(X + Y) = E(X) + E(Y)
  - ☆ In general, E(XY) ≠ E(X)E(Y); If X and Y are independent random variables, then E(XY) = E(X)E(Y).

二、Variance 方差：

- **Variance** 1.定义：衡量数据的离散程度.
  - A measure of dispersion – the second moment.

2.公式：

$$\boxed{\sigma^2 = E(X-\mu)^2}$$ 定义式.

  - Above formula is the definition of variance. To compute the variance, we use the following formula:

$$\boxed{\sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2}$$ 计算式.

  - Measures how noisy or unpredictable that random variable is.
  - The positive square root of σ_x^2, σ_X, is known as the standard deviation, also called volatility.

$$\sigma^2 = E(x-\mu)^2$$

$$E(X+Y) = E(X) + E(Y)$$

$$= E(x^2 + \underline{\mu}^2 - 2x \cdot \underline{\mu})$$

$$= E(x^2 - 2x \cdot \mu) + \mu^2$$

$$= \underline{E(x^2) - 2E(x \cdot \mu)} + \mu^2$$

GOLDEN FUTURE

$$\sigma^2 = E(x-\mu)^2$$

$$E(X+Y) = E(X) + E(Y)$$

$$= E(X^2 + \mu^2 - 2X\cdot\mu)$$

$$= E(X^2 - 2X\cdot\mu) + \mu^2 \qquad \mu = E(X)$$

$$= E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - 2\cdot E(X)\cdot E(X) + [E(X)]^2$$

$$\sigma^2 = E(X-\mu)^2 \text{ 定义式}$$

$$E(X+Y) = E(X) + E(Y)$$

$$= E(X^2 + \mu^2 - 2X\cdot\mu)$$

$$= E(X^2 - 2X\cdot\mu) + \mu^2 \qquad \mu = E(X)$$

$$= E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - 2[E(X)]^2 + [E(X)]^2$$

$$\boxed{= E(X^2) - [E(X)]^2} \text{ 计算式}$$

Variance 方差：

- **Variance** 1.定义：衡量数据的离散程度.
  - A measure of dispersion – the second moment.

2.公式：

$$\boxed{\sigma^2 = E(X-\mu)^2.}$$

定义式.

  - Above formula is the definition of variance. To compute the variance, we use the following formula:

$$\boxed{\sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2}$$

计算式（推导）

  - Measures how noisy or unpredictable that random variable is.
  - The positive square root of σ_x^2, σ_X, is known as the standard deviation, also called volatility. 波动率

标准差 1 2

variance.

阶矩.

Variance

3.性质

> **Properties of Variance**

- If c is constant, then: $\sigma^2(c) = 0$.
- If a is constant, then: $\sigma^2(aX) = a^2\sigma^2(X)$.
- If b is a constant, then: $\sigma^2(X + b) = \sigma^2(X)$.
- If a and b are constant, then: $\sigma^2(aX + b) = a^2\sigma^2(X)$.
- If X and Y are independent random variables and a and b are constants, then $\sigma^2(aX + bY) = a^2\sigma^2(X) + b^2\sigma^2(Y)$.
- The relationship between expectation and variance:

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

Sample Mean

➢ **Sample Mean**

- The sample mean of a random variable, X, is defined as:

$$\boxed{\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}}$$

算术

- The sample mean is known as an estimator of E(X), which we can now call the population mean.
- An estimate of the population is simply the numerical value taken by an estimator.

Sample Variance

Sample Variance

- The sample variance, denoted by S_x^2 which is an estimator of σ_x^2, which we can now call the population variance. The sample variance is defined as:

$$\boxed{S_x^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}}$$

- The expression (n – 1) is known as the degrees of freedom.
- If the sample size is reasonably large, we can divide by n instead of (n – 1).
- S_X (the positive square root of S_x^2), is called the sample standard deviation.

Sample Mean

➢ Sample Mean

- The sample mean of a random variable, X, is defined as:

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

算术

估计量

μ

- The sample mean is known as an estimator of E(X), which we can now call the population mean.
- An estimate of the population is simply the numerical value taken by an estimator

估计量：用来估计总体

Sample Variance

➢ Sample Variance

- The sample variance, denoted by S_x^2 which is an estimator of σ_x^2, which we can now call the population variance. The sample variance is defined as:

$E(x-\mu)^2$ 自由度：可以自由变动的量

$$\boxed{S_x^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}}$$

- The expression (n – 1) is known as the degrees of freedom.
- If the sample size is reasonably large, we can divide by n instead of (n – 1).
- S_X (the positive square root of S_x^2), is called the sample standard deviation.

Sample Mean and Variance

	Population	Sample
Mean	$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$
Standard Deviation	σ	s

Best Linear Unbiased Estimator (BLUE) 最小线性无偏估计量

> BLUE：衡量估计量好坏的标准，

- Another property of a point estimate is linearity. A point estimate should be a linear estimator (i.e., it can be used as a linear function of the sample data). If the estimator is the best available (i.e., has the minimum variance), exhibits linearity, and is unbiased, it is said to be the best linear unbiased estimator (BLUE).
- All of the estimators that we produced in this chapter for the mean, variance, covariance, skewness, and kurtosis are either BLUE or the ratio of BLUE estimators.

Best Linear Unbiased Estimator (BLUE) 最小残差 无偏估计量

- BLUE：衡量估计量好坏的标准。

unbaised
effecient
consistency
一致性
n个
误差↓

- Another property of a point estimate is linearity. A point estimate should be a linear estimator (i.e., it can be used as a linear function of the sample data). If the estimator is the best available (i.e., has the minimum variance), exhibits linearity, and is unbiased, it is said to be the best linear unbiased estimator (BLUE).

无偏性：$E(\bar{X}_1) = \mu$ $E(\bar{X}_1 \quad \bar{X}_2 \quad \bar{X}_3 \; \bar{X}_4$

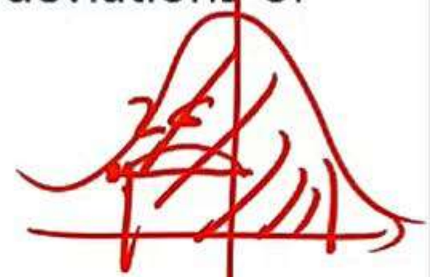
- All of the estimators that we produced in this chapter for the mean, variance, covariance, skewness, and kurtosis are either BLUE or the ratio of BLUE estimators.

$\sqrt{}$

Chebyshev's Inequality

- **Chebyshev's Inequality** 契比雪夫不等式：
  - 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 $1 - 1/k^2$, $k > 1$.

$$\boxed{P(|X-\mu| \le k\sigma) \ge 1-\frac{1}{k^2} \quad k>1}$$



  - This relationship applies regardless of the shape of the distribution.

Lin within	2	Standard deviations of the mean	$\ge 1-\frac{1}{2^2} = 1-\frac{1}{4} = \frac{3}{4} = 75\%$
	3		$\ge 1-\frac{1}{3^2} = 1-\frac{1}{9} = \frac{8}{9} = 89\%$
	4		$\ge 1-\frac{1}{4^2} = 1-\frac{1}{16} = \frac{15}{16} = 94\%$

Chebyshev's Inequality

- **Chebyshev's Inequality** 切比雪夫不等式：

  对所有分布来讲

  - 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 $1 - 1/k^2$, $k > 1$.

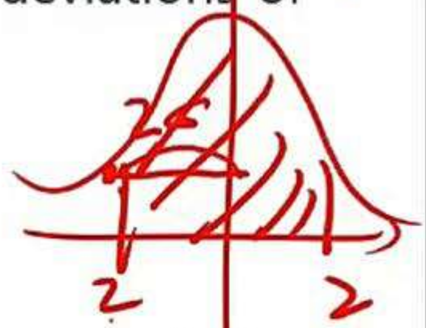
95%

25%

2.

$$\boxed{P(|X-\mu| \le k\sigma) \ge 1 - \frac{1}{k^2}\quad k > 1}$$

10 ≤ 2



- This relationship applies regardless of the shape of the distribution.

$1-\frac{1}{5^2}$ =96%

3 $\mu=170$ $\sigma=2$

[160, 180]

求prob.

Lin within	2	Standard deviations of the mean	$\ge 1-\frac{1}{2^2}=1-\frac{1}{4}=\frac{3}{4}=75\%$
	3		$\ge 1-\frac{1}{3^2}=1-\frac{1}{9}=\frac{8}{9}=89\%$
	4		$\ge 1-\frac{1}{4^2}=1-\frac{1}{16}=\frac{15}{16}=94\%$

Covariance 协方差：两组数据的关系

$E[x - E(x)]^2$

$E(x^2) - [E(x)]^2$

- **Covariance**

$$Cov(X,Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] = E(XY) - E(X)E(Y)$$

- Covariance measures how one random variable moves with another random variable.
- Covariance ranges from negative infinity to positive infinity.

- **Properties of Covariance**

- If X and Y are independent random variables, their covariance is zero.

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

$$Cov(a + bX,\ cY) = Cov(a, cY) + Cov(bX, cY) = b \times c \times Cov(X,Y)$$

- The relationship between covariance and variance:

$\sigma2(X \pm Y) = \sigma2(X) + \sigma2(Y) \pm 2Cov(X,Y)$

2. $cov(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$

$$
cov\begin{cases} >0 & X>\mu_X \quad Y>\mu_Y \\ =0 \\ <0 \end{cases}
$$

正向关系、

没有关系

负向关系

Covariance

协方差：两组数据的关系

$E[X - E(X)]^2$

$E(X^2) - [E(X)]^2$

- **Covariance**

$$\boxed{Cov(X,Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] = E(XY) - E(X)E(Y)}$$

  - Covariance measures how one random variable moves with another random variable.
  - Covariance ranges from negative infinity to positive infinity.

3

- **Properties of Covariance**
  - If X and Y are independent random variables, their covariance is zero.

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

$$Cov(a + bX,\ cY) = \cancel{Cov(a, cY)} + Cov(bX, cY) = b \times c \times Cov(X,Y)$$

  - The relationship between covariance and variance:

σ^2

σ2(X ± Y) = σ2(X) + σ2(Y) ± 2Cov(X,Y)

4. 应用人 美

$$\mathrm{cov} = E\underbrace{(X-\mu_X)}_{RMB}\cdot\underbrace{(Y-\mu_Y)}_{USD}$$

$$\mathrm{cov}(\overset{人}{X}, \overset{美}{Y}) = 100\ \mathrm{RMB}\cdot\mathrm{USD}$$

$$\mathrm{cov}(\underset{人}{X}, \underset{日}{Z}) = +1000 : \mathrm{RMB}\cdot\mathrm{JPY}$$

$$\mathrm{cov}(Y, Z) =$$

~~斤~~ / kg　km

Correlation Coefficient

-1 0 1

3. Correlation coefficient

① [-1, 1]

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

$\bar{X} \cdot \bar{Y}$

② Properties of Correlation coefficient

- Correlation has no units, ranges from -1 to +1.
- Correlation measures the linear relationship between two random variables.

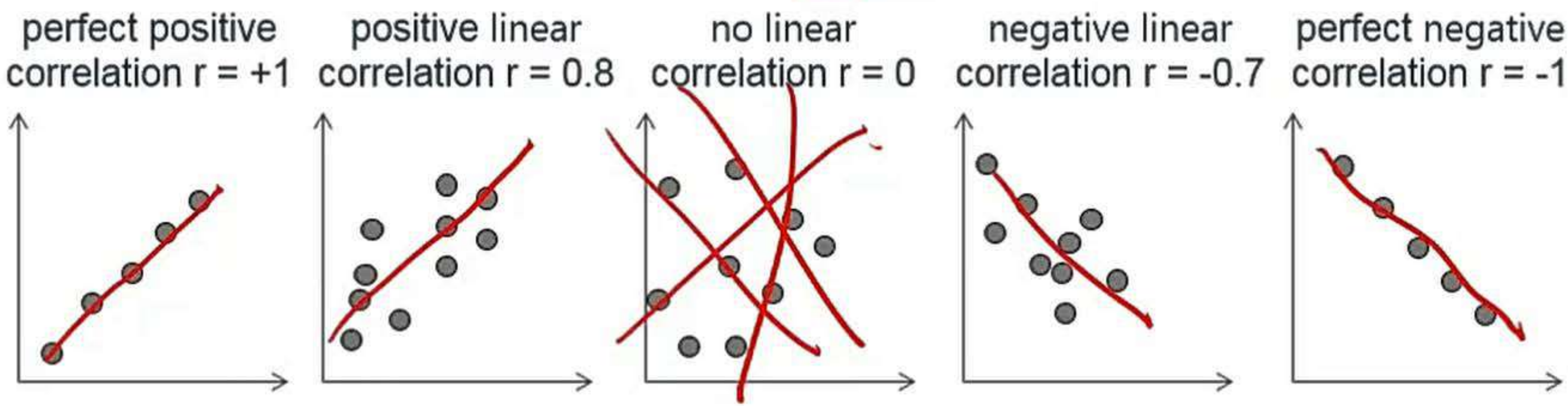
线性关系

- If two variables are independent, their covariance is zero, therefore, the correlation coefficient will be zero. The converse, however, is not true. For example, $Y = X^2$.
- Variances of correlated variables:

$$\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y) \pm 2\rho\sigma(X)\sigma(Y)$$

Correlation Coefficient

Correlation coefficient	Interpretation	
ρ $r = +1$	perfect positive correlation	完全正相关
$0 < r < +1$	positive linear correlation	正相关
$r = 0$	no linear correlation	没有线性关系
$-1 < r < 0$	negative linear correlation	负相关
$r = -1$	perfect negative correlation	完全负相关



GOLDEN FUTURE

Measures of Portfolio

$$E(R_P) = \sum_{i=1}^{N} w_i E(R_i) = w_1 E(R_1) + w_2 E(R_2) + \cdots + w_n E(R_N)$$

$$\sigma^2(R_P) = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j Cov(R_i, R_j)$$

$$w_i = \frac{market\ value\ of\ investment\ in\ asset\ i}{market\ value\ of\ the\ portfolio}$$

2个. $\sigma_p^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$.

3个 $\sigma_p^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\rho_{1,2}\sigma_1\sigma_2 + 2\rho_{1,3}\sigma_1\sigma_3 + 2\rho_{2,3}\sigma_2\sigma_3$

Measures of Portfolio

$$E(R_P) = \sum_{i=1}^{N} w_i E(R_i) = w_1 E(R_1) + w_2 E(R_2) + \cdots + w_n E(R_N)$$

$$\sigma^2(R_P) = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j Cov(R_i, R_j)$$

$$w_i = \frac{market\ value\ of\ investment\ in\ asset\ i}{market\ value\ of\ the\ portfolio}$$

特殊：

perfect：$\rho \pm 1$

$\sigma_p = \sigma_1 + \sigma_2$

$\rho = -1$

$\sigma_p = |\sigma_1 - \sigma_2|$

2个：$\sigma_p^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$

3个：$\sigma_p^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\rho_{1,2}\sigma_1\sigma_2 + 2\rho_{1,3}\sigma_1\sigma_3 + 2\rho_{2,3}\sigma_2\sigma_3$

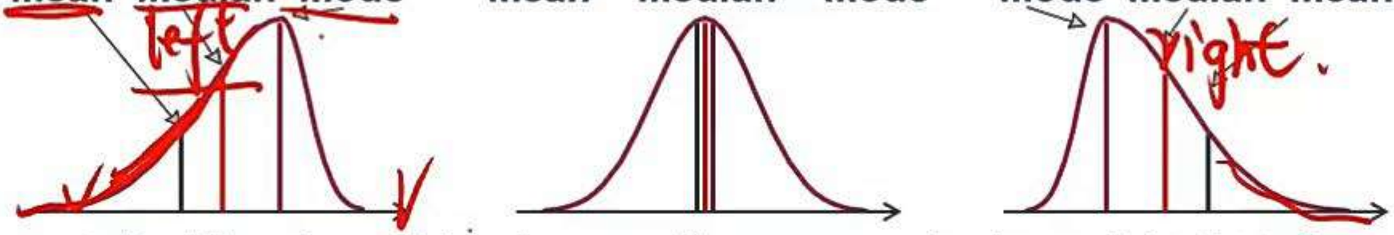
三、Skewness 偏度

- **Skewness** 1、衡量数据对称性的.
  - A measure of asymmetry of a PDF – the third moment.

2. $$\boxed{S = \frac{E(X-\mu_x)^3}{\sigma_x^3} = \frac{\text{third moment about mean}}{\text{cube of standard deviation}}}$$ 不要求计算

3. - **Symmetrical and nonsymmetrical distributions**
  - Positively skewed (right skewed) and negatively skewed(left skewed)

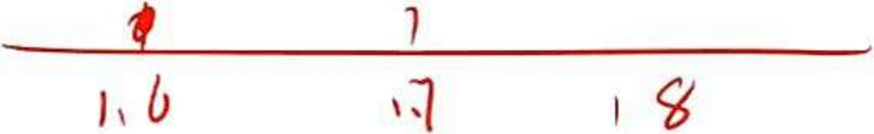
Negative-skewed	Symmetric	Positive-skewed
Mean Median Mode	Mean = Median = Mode	Mode Median Mean



- Positive skewed: Mode < median < mean, having a right fat tail
- Negative skewed: Mode > media > mean, having a left fat tail

mean	median	mode
1.6	1.7	1.8

positive / <u>negative</u>



四 Kurtosis 峰度

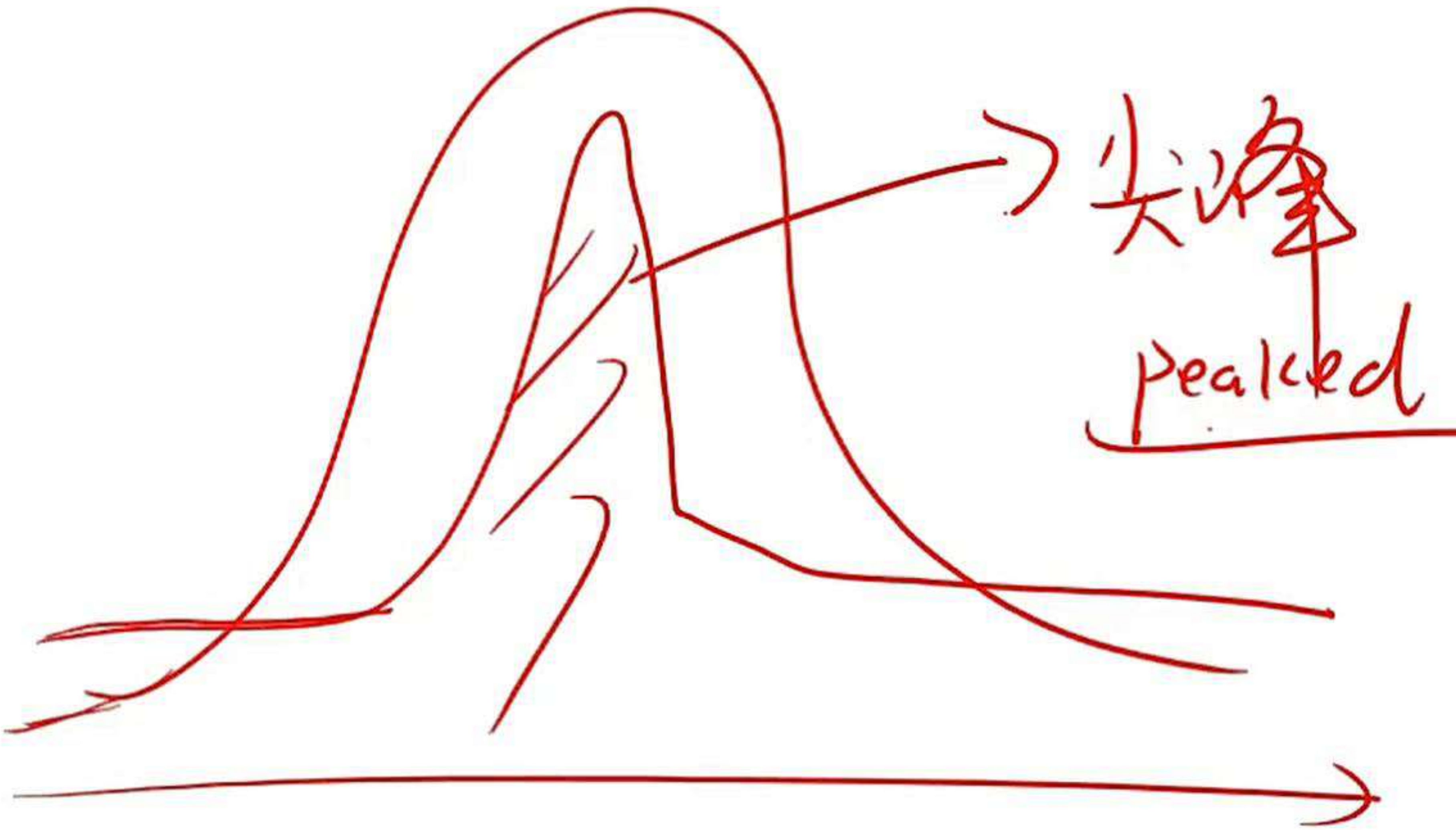
- **Kurtosis** ：1. 衡量一组数据陡峭程度 / 尾巴薄厚
  - A measure of tallness or flatness of a PDF – the fourth moment.

$$\boxed{K = \frac{E(X-\mu_x)^4}{[E(X-\mu_x)^2]^2} = \frac{\text{fourth moment}}{\text{square of second moment}}}$$

  - For a normal distribution, the K value is 3.
  - Excess kurtosis = kurtosis – 3

12

	leptokurtic	mesokurtic	platykurtic
Kurtosis	> 3	= 3	< 3
Excess kurtosis	> 0	= 0	< 0
Tails (assuming same variance)	fat tail	normal	thin tail


尖峰
peaked

四 Kurtosis 峰度

➢ **Kurtosis**：1.衡量一组数据陡峭程度 / 尾巴薄厚

- A measure of tallness or flatness of a PDF – the fourth moment.

$$\boxed{K = \frac{E(X-\mu_x)^4}{[E(X-\mu_x)^2]^2} = \frac{\text{fourth moment}}{\text{square of second moment}}}$$

方差一致

- For a normal distribution, the K value is 3.
- Excess kurtosis = kurtosis – 3

尖峰肥厚
矮峰瘦尾

	leptokurtic	mesokurtic	platykurtic
Kurtosis	> 3	= 3	< 3
Excess kurtosis	> 0	= 0	< 0
Tails (assuming same variance)	fat tail	normal	thin tail

四 Kurtosis 峰度

➢ **Kurtosis**：1. 衡量一组数据陡峭程度 / 尾巴薄厚

- A measure of tallness or flatness of a PDF – the fourth moment.

$$\boxed{K = \frac{E(X-\mu_x)^4}{[E(X-\mu_x)^2]^2} = \frac{\text{fourth moment}}{\text{square of second moment}}}$$

方差一致

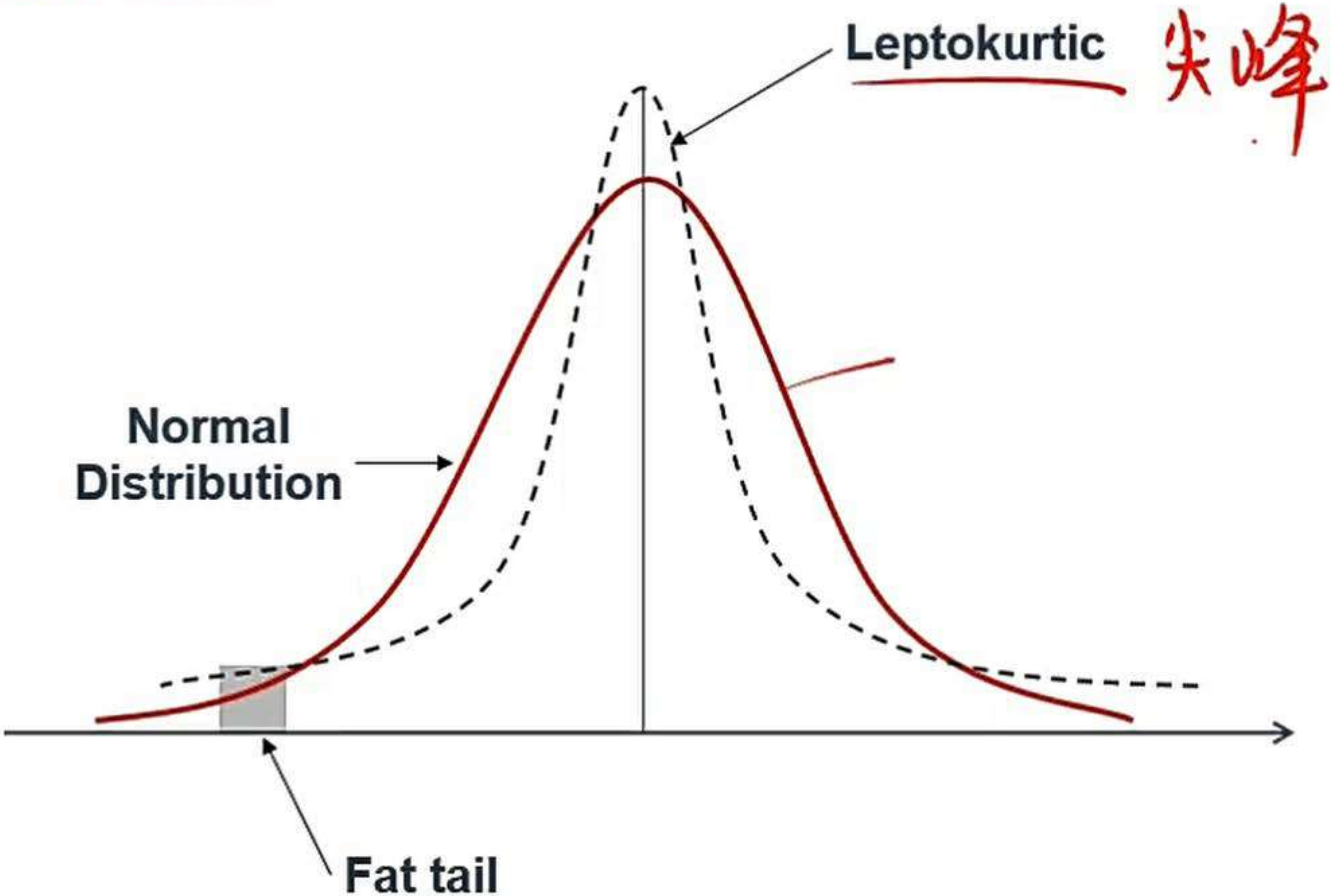
尖峰肥厚

矮峰瘦尾

- For a normal distribution, the K value is 3.
- Excess kurtosis = kurtosis – 3

	尖峰 leptokurtic	mesokurtic	矮 platykurtic
Kurtosis	> 3	= 3	< 3
Excess kurtosis	> 0	= 0	< 0
Tails (assuming same variance)	fat tail	normal	thin tail

Kurtosis



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

Coskewness and Cokurtosis

- Just as we generalized the concept of mean and variance to moments and central moments, we can generalize the concept of covariance to **cross central moments.**
- **Coskewness and Cokurtosis**
  - The third cross central moment is referred to as **coskewness**.
  - The fourth cross central moment is referred to as **cokurtosis**.

Coskewness and Cokurtosis

- The reason the above charts look different or the reason the returns of the two portfolios are different, is because the coskewness between the portfolios is different.

	A and B	C and D
S_{xxy}	0.99	-0.58
S_{xyy}	0.58	-0.99

➢ **Notices**

- The nontrivial coskewness of two variables: S_{XXY} and S_{XYY}
  - ✓ For example $$\boxed{S_{XXY} = \frac{E[(X-\mu_x)^2(Y-\mu_Y)]}{\sigma_X^2\sigma_Y}}$$ S_{XYY}
- The nontrivial cokurtosis of two variables: K_{XXXY}、K_{XXYY} and K_{XYYY}
  - ✓ For example $$\boxed{K_{XXXY} = \frac{E[(X-\mu_x)^3(Y-\mu_Y)]}{\sigma_X^3\sigma_Y}}$$ K_{XXYY} K_{XYYY}

Coskewness and Cokurtosis

➢ Example

- Assume four series of fund returns (A、B、C、D) where the mean, standard deviation, skew, and kurtosis are all the same, but only the order of returns is different:

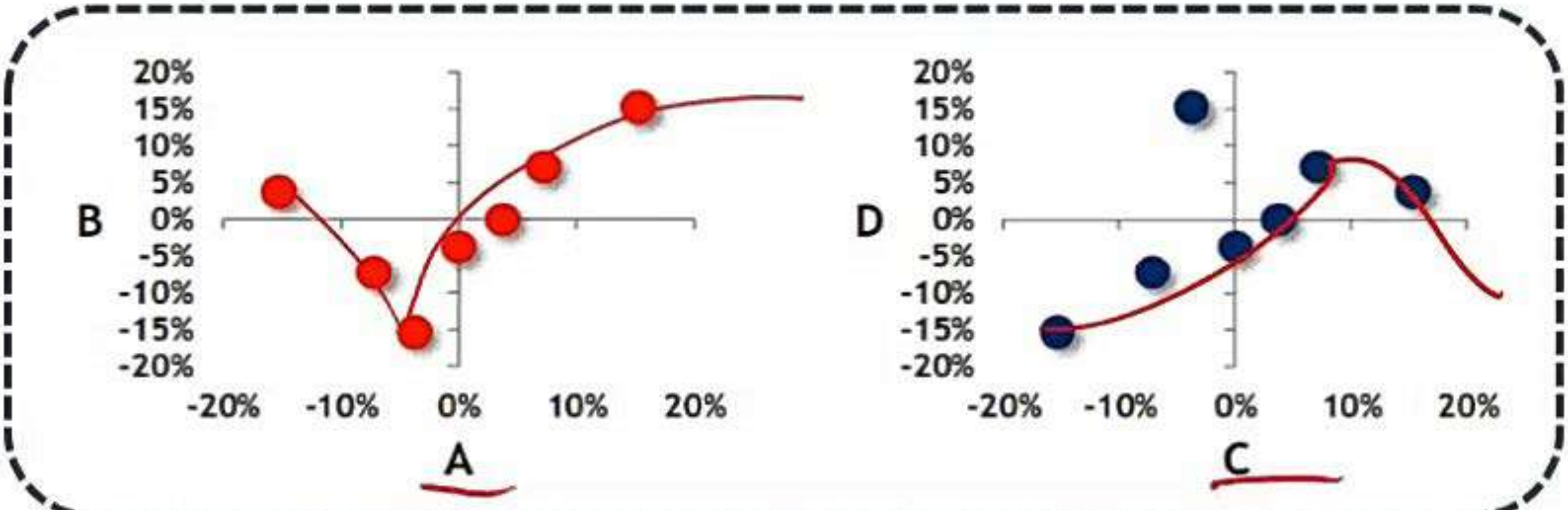
time	A	B	C	D
1	0.0%	-3.8%	-15.3%	-15.3%
2	-3.8%	-15.3%	-7.2%	-7.2%
3	-15.3%	3.8%	0.0%	-3.8%
4	-7.2%	-7.2%	-3.8%	15.3%
5	3.8%	0.0%	3.8%	0.0%
6	7.2%	7.2%	7.2%	7.2%
7	15.3%	15.3%	15.3%	3.8%

time	A + B	C + D
1	-1.9%	-15.3%
2	-9.5%	-7.2%
3	-5.8%	-1.9%
4	-7.2%	5.8%
5	1.9%	1.9%
6	7.2%	7.2%
7	15.3%	9.5%

- The two portfolios (A + B and C + D) have the same mean and standard deviation, but the skews of the portfolios are different.

Coskewness and Cokurtosis

- Scatterplots show the difference between B versus A and D versus C:
  - ✓ A and B: their best positive returns occur during the same time period, but their worst negative returns occur in different periods. This causes the distribution of points to be skewed toward the top-right of the chart.
  - ✓ C and D: their worst negative returns occur in the same period, but their best positive returns occur in different periods. In the second chart, the points are skewed toward the bottom-left of the chart.



Example 1

B

- Which one of the following statements about the correlation coefficient is false?

  ρ

  A. It always ranges from -1 to +1.

  B. A correlation coefficient of zero means that two random variables are independent. $\rho = 0$ $\Rightarrow$ indopendant

  C. It is a measure of linear relationship between two random variables

  D. It can be calculated by scaling the covariance between two random variables.

- Correct Answer: B

$$\operatorname{cov}(ax+by,\ cx+dy)$$

$$=\operatorname{cov}(ax,\ cx+dy)+\operatorname{cov}(by,\ cx+dy)$$

$$=\operatorname{cov}(ax,\ cx)+\operatorname{cov}(ax,\ dy)+\operatorname{cov}(by,\ cx)$$

$$\operatorname{cov}(by,\ dy)$$

$$=ac\,\sigma_x^2+\underline{ad\operatorname{cov}(x,Y)}+\underline{bc\operatorname{cov}(x,Y)}$$

$$+\,bd\,\sigma_Y^2$$

$$=\underline{ac\,\sigma_x^2+bd\,\sigma_Y^2+(ad+bc)\operatorname{cov}(X,Y)}$$

Example 2

- Given that x and y are random variables, and a, b, c and d are constant, which one of the following definitions is wrong?
   A. $E(ax + by + c) = aE(x) + bE(y) + c$, if x and y are correlated.
   B. $\sigma^2(ax + by + c) = \sigma^2(ax + by) + c$, if x and y are correlated.
   C. $Cov(ax + by, cx + dy) = ac\sigma^2(x) + bd\sigma^2(y) + (ad + bc)Cov(x, y)$, ~~if x and y are correlated.~~
   D. $\sigma^2(x - y) = \sigma^2(x + y) = \sigma^2(x) + \sigma^2(y)$, if x and y are uncorrelated.

- Correct Answer: B

$$\sigma^2(x+b) = \sigma^2(x)$$

Example 2

- Given that x and y are random variables, and a, b, c and d are constant, which one of the following definitions is wrong?
  - A. $E(ax + by + c) = aE(x) + bE(y) + c$, if x and y are correlated.
  - B. $\sigma^2(ax + by + c) = \sigma^2(ax + by) + c$, if x and y are correlated.
  - C. $Cov(ax + by, cx + dy) = ac\sigma^2(x) + bd\sigma^2(y) + (ad + bc)Cov(x, y)$, if x and y are correlated.
  - D. $\sigma^2(x - y) = \sigma^2(x + y) = \sigma^2(x) + \sigma^2(y)$, if x and y are uncorrelated.

$-2cov(x,y)$ $+2cov(x,y)$

$\sigma^2(x+b) = \sigma^2(x)$

- Correct Answer: B

总结：

一、阶矩：mean

① 衡量中心趋势

②. 公式

③ 性质

④ sample mean

GOLDEN FUTURE

二阶矩：

1. variance ①衡量离散程度

②公式：定义、计算

③性质：

④ sample variance

2. covariance：① 两组数据的相关性.

② 公式：定义、计算

③ 性质：比正负、但不能比大小

3 相关系数：① 线性关系

② 公式

③ 性质

12

三阶矩：偏度 ① 衡量对称性.

② 性质： mean median mode

四阶矩：峰度 ① 衡量陡峭程度/尾巴薄厚

② 性质

五. BLUE：衡量估计量的好坏标准

六. Chebychev's：公式

2个离散
- 二项
- 泊松

6个连续
- 连续均匀
- normal
- lognormal
- chi-square
- T-distribution
- F-distribution

Distributions

八个分布

Binomial Distribution 二项分布

- Bernoulli Distribution 伯努力分布.
- P(X = 1) = p   P(X = 0) = 1 − p

➢ Binomial Distribution

- The probability of x successes in n trails

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

➢ Expectations and variances

	Expectation	Variance
Bernoulli random variable	p	p(1 − p)
Binomial random variable	np	np(1 − p)

硬币 正（成）p 反（失） 1次

成功次数	prob
0^2	$1-p$
1^2	p

$$\text{mean} = 0 \times (1-p) + 1 \times p = p$$

$$\text{variance} = E(x^2) - [E(x)]^2$$

$$= \cancel{0^2 \times (1-p)} + 1^2 \times p - p^2$$

$$= p - p^2 = p(1-p)$$

Binomial Distribution 二项分布

- Bernoulli Distribution 伯努力分布.
- $P(X = 1) = p \quad P(X = 0) = 1 - p$

> **Binomial Distribution** : 独立重复n次伯努力实验.

- The probability of x successes in n trails

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

> **Expectations and variances**

	Expectation	Variance
Bernoulli random variable	p	p(1 – p)
Binomial random variable	np	np(1 – p)

硬币 10次 P

成功次数 prob

0 $C_{10}^{0}(1-P)^{10}P^{0}$

1 $C_{10}^{1}P(1-P)^{9}$

2 $C_{10}^{2}P^{2}(1-P)^{8}$

3

⋮

10 $C_{10}^{10}P^{10}(1-P)^{0}$

mean = nP

Variance =

$n p(1-P)$

$\sigma_p^2=\sigma_{(1+2)}^2=\sigma_1^2+\sigma_2^2$

$\cancel{+2\rho\cdot\sigma_1\sigma_2}$

$2P(1-P)$

一、Binomial Distribution 二项分布

- Bernoulli Distribution 伯努利分布.
- P(X = 1) = p P(X = 0) = 1 − p .

➢ **Binomial Distribution** 定义: 独立重复n次伯努利实验.

- The probability of x successes in n trails

2、公式

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

➢ **Expectations and variances**

3. 性质

	Expectation	Variance
Bernoulli random variable	p	p(1 − p)
Binomial random variable	np	np(1 − p)

Some Important Probability Distributions

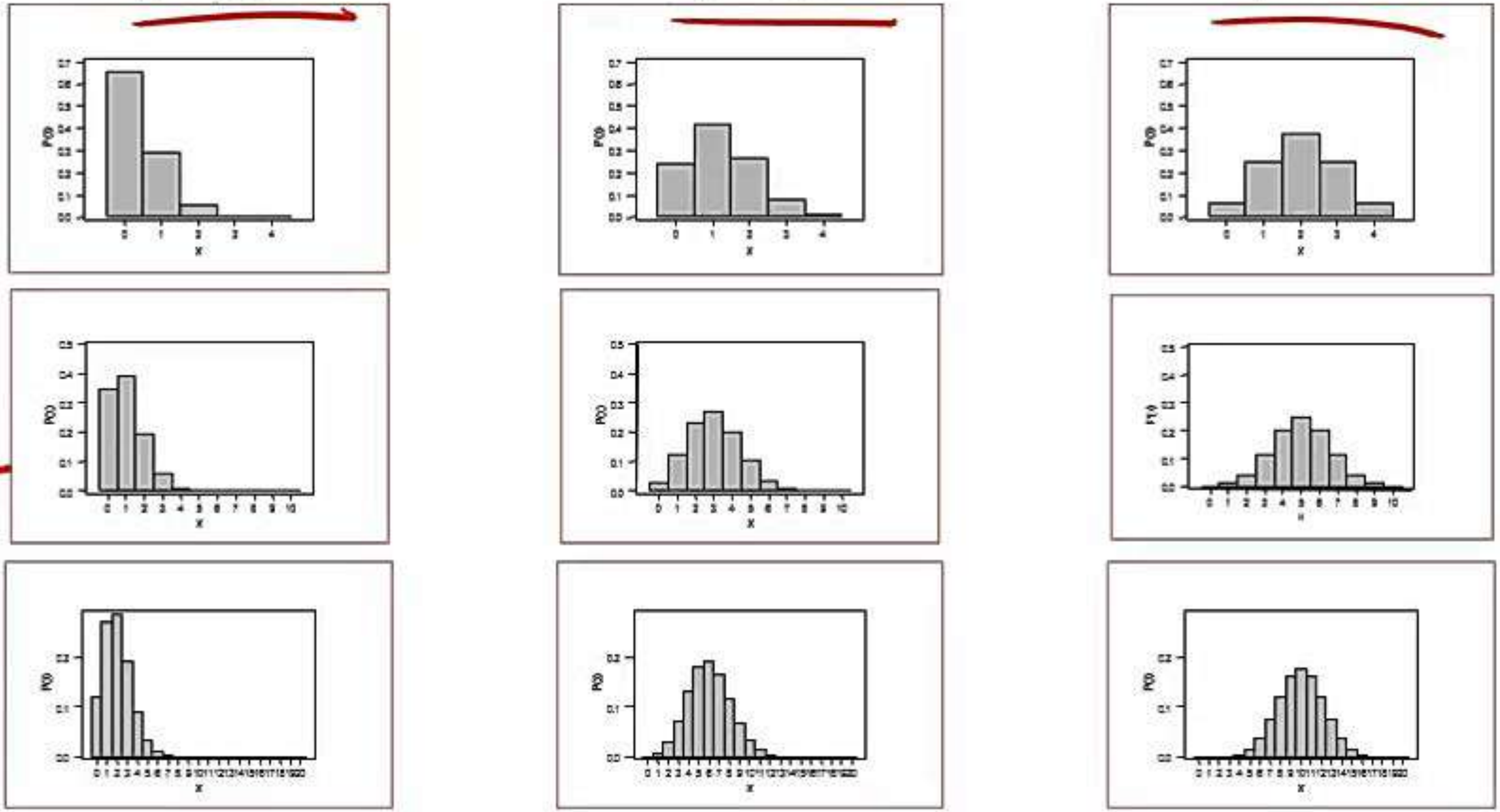
- The Binomial Distribution – Overview

p = 0.1 | p = 0.3 | p = 0.5

n = 4

n = 10

n = 20



- Binomial distributions become more symmetric as $p = 0.5$.

Poisson Distribution

- **Poisson Distribution** ：n极大 p极小、 np = constant
  - When there are a large number of trials but a small probability of success, Binomial calculations become impractical.
  - If we substitute λ/n for p, and let n very large, the Binomial Distribution becomes the Poisson Distribution.

$$\boxed{p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \ \ (\lambda = np)}$$

✓ X refers to the number of success per unit.

✓ λ indicates the rate of occurrence of the random events; i.e., it tells us how many events occur on average per unit of time.

- Example:

✓ The number of fish caught in a day; the number of potholes on a 1 km stretch of road; the number of persons appeared in a shopping mall; the number of phone calls in a day.

二项 $\longrightarrow$ possion. $n \to \infty \quad p \to 0$.

$$\frac{n!}{x!\,\cancel{(n-x)!}} \cdot p^x (1-p)^{n-x}$$

$$\lim_{n\to\infty,\, p\to 0} = \frac{n \times (n-1) \cdots \times (n-x+1)}{x!} p^x (1-p)^{n-x}$$

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

二项 $\longrightarrow$ possion. $n \to \infty$ $p \to 0$.

$$\frac{n!}{x!\cancel{(n-x)!}} \cdot p^x (1-p)^{n-x}$$

$$\lim_{n\to\infty,\, p\to 0} = \frac{n\times(n-1)\cdots\times(n-x+1)}{x!} p^x (1-p)^{n-x}$$

$$= \boxed{\frac{n^x p^x}{x!}(1-p)^n} \cdot \cancel{\frac{1}{(1-p)^x}}$$

$$= \frac{(np)^x}{x!}$$

二项 $\longrightarrow$ possion. $n \to \infty$ $p \to 0$.

$$\frac{n!}{x!\,\cancel{(n-x)}!} \cdot p^x (1-p)^{n-x} \qquad np =$$

$$\lim_{n\to\infty,\, p\to 0} = \frac{\underline{n \times (n-1)} \cdots \times \underline{(n-x+1)}}{x!} p^x (1-p)^{n-x}$$

$$= \boxed{\frac{n^x p^x}{x!}(1-p)^n} \cdot \cancel{\frac{1}{(1-p)^x}} \qquad \lambda = np$$

$$= \frac{\lambda^x}{x!}(1-p)^{\frac{\lambda}{p}}.$$

12

二项 $\longrightarrow$ possion. $n \to \infty$ $p \to 0$.

$$\frac{n!}{x!\,\cancel{(n-x)!}} \cdot p^x (1-p)^{n-x} \qquad p=$$

$$\lim_{n\to\infty,\, p\to 0} = \frac{n \times (n-1) \cdots \times (n-x+1)}{x!} p^x (1-p)^{n-x}$$

$$= \boxed{\frac{n^x p^x}{x!} (1-p)^n} \cdot \cancel{\frac{1}{(1-p)^x}} \qquad \lambda = np$$

$$= \boxed{\frac{\lambda^x}{x!}}\, \cancel{(1-p)^{\frac{\lambda}{p}}} \cdot \frac{\boxed{(1-p)^{-\frac{1}{p}}}}{e} \cdot -\lambda$$

e $= 1$

12

Poisson Distribution

- **Poisson Distribution** ：n极大 p极小、 np = constant
  - When there are a large number of trials but a small probability of success, Binomial calculations become impractical.
  - If we substitute λ/n for p, and let n very large, the Binomial Distribution becomes the Poisson Distribution. λ、单位时间某件事发生的平均次数.

2、 $$\boxed{p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \ (\lambda = np)}$$

  - ✓ X refers to the number of success per unit.
  - ✓ λ indicates the rate of occurrence of the random events; i.e., it tells us how many events occur on average per unit of time.
- Example:
  - ✓ The number of fish caught in a day; the number of potholes on a 1 km stretch of road; the number of persons appeared in a shopping mall; the number of phone calls in a day.

Some Important Probability Distributions

- **Example**
  - A company receives three complaints per day on average. What is the probability of receiving more than one complaint on a particular day?

5个

10天内

  - $\lambda = 3$
    - ✓ "more than one" means that k = 2 or 3 or 4 or ...
    - ✓ P('more than one') = P(2) + P(3) + P(4) + ...
    - ✓ P('more than one') = 1 – {P(0) + P(1)}
    - ✓ $P(0) = e^{-3} \times 3^0 / 0! = 0.0498$
    - ✓ $P(1) = e^{-3} \times 3^1 / 1! = 0.1494$
    - ✓ P(0) + P(1) = 0.1992
    - ✓ P('more than one') = 1 – {P(0) + P(1)} = 1 – 0.1992 = 0.8008

Poisson Distribution

3.性质 方差

$E(x) = np = \lambda$

$V(x) = np(1-p) = \lambda \times 1$

➤ **Properties**

- $E(X) = D(X) = \lambda$
- The sum of independent Poisson variables is a further Poisson variable with mean equal to the sum of the individual means.
- The Poisson Distribution is the limiting case of the Binomial Distribution as n goes to infinity and p goes to zero, while np = λ remains fixed. In addition, when λ is large the Poisson Distribution is well approximated by the Normal Distribution with mean and variance of λ, through the central limit theorem.

Some Important Probability Distributions

泊松. $\frac{\lambda^k}{k!}e^{-\lambda}$

- **Example**
  - A company receives three complaints per day on average. What is the probability of receiving more than one complaint on a particular day?

    5个 10天内

  - $\lambda = 3$
    - ✓ "more than one" means that k = 2 or 3 or 4 or … >1
    - ✓ P('more than one') = P(2) + P(3) + P(4) + …
    - ✓ P('more than one') = 1 – {P(0) + P(1)}
    - ✓ P(0) = $e^{-3} \times 3^0 / 0!$ = 0.0498 +
    - ✓ P(1) = $e^{-3} \times 3^1 / 1!$ = 0.1494
    - ✓ P(0) + P(1) = 0.1992
    - ✓ P('more than one') = 1 – {P(0) + P(1)} = 1 – 0.1992 = 0.8008

k

1个 $\frac{3^1}{1!}e^{-3}$

0个 $\frac{3^0}{0!}e^{-3}$

12

Example 2

- A call center receives an average of two phone calls per hour. The probability that they will receive 20 calls in an 8-hour day is closest to:

A. 5.59%

B. 16.56%

C. 3.66%

D. 6.40%

$$\frac{\lambda^K}{K!} e^{-\lambda} \qquad K = 20 \qquad \lambda = 2 \times 8$$

- Correct Answer : A
  - To solve this question, we first need to realize that the expected number of phone calls in an 8-hour day is 16. Using the Poisson distribution, we solve for the probability that X will be 20.
  - $$P(X = 20) = \frac{16^{20} e^{-16}}{20!} = 5.59\%$$

Continuous Uniform Distribution

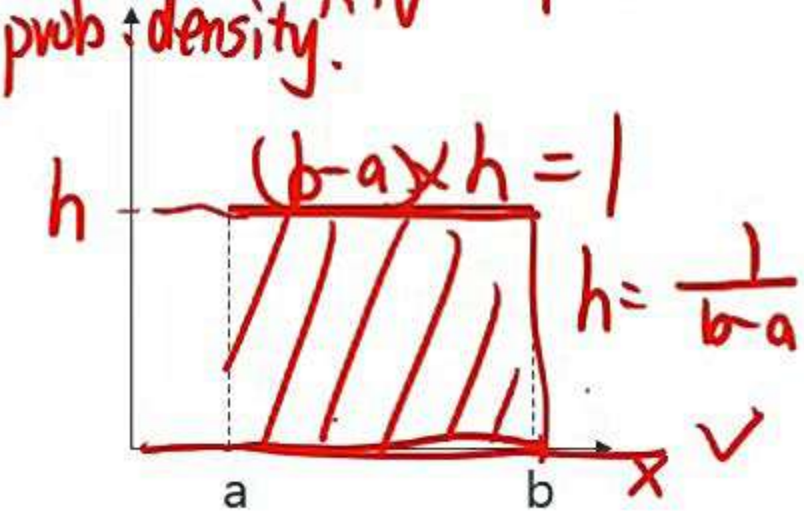
连续 均匀

R.V prob

Probability density function

函数

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



Cumulative distribution function

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

PF / PDF

eg: 3:03 − 3:04

$\boxed{3:03:25' - 3:03:45'} = \frac{1}{3}$

Continuous Uniform Distribution

> **Properties**

- $E(X) = (a + b)/2$
- $D(X) = (b - a)^2/12$
- For all $a \le x_1 < x_2 \le b$, we have:

$$\boxed{P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx = \frac{x_2 - x_1}{b - a}}$$

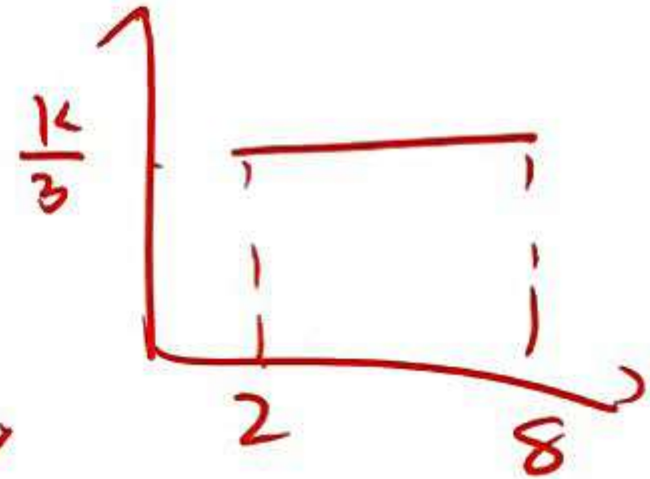
> **Example**

- The random variable X with density function $f(x) = k/3$ for $2 \le x \le 8$, and 0 otherwise. Calculate its mean.

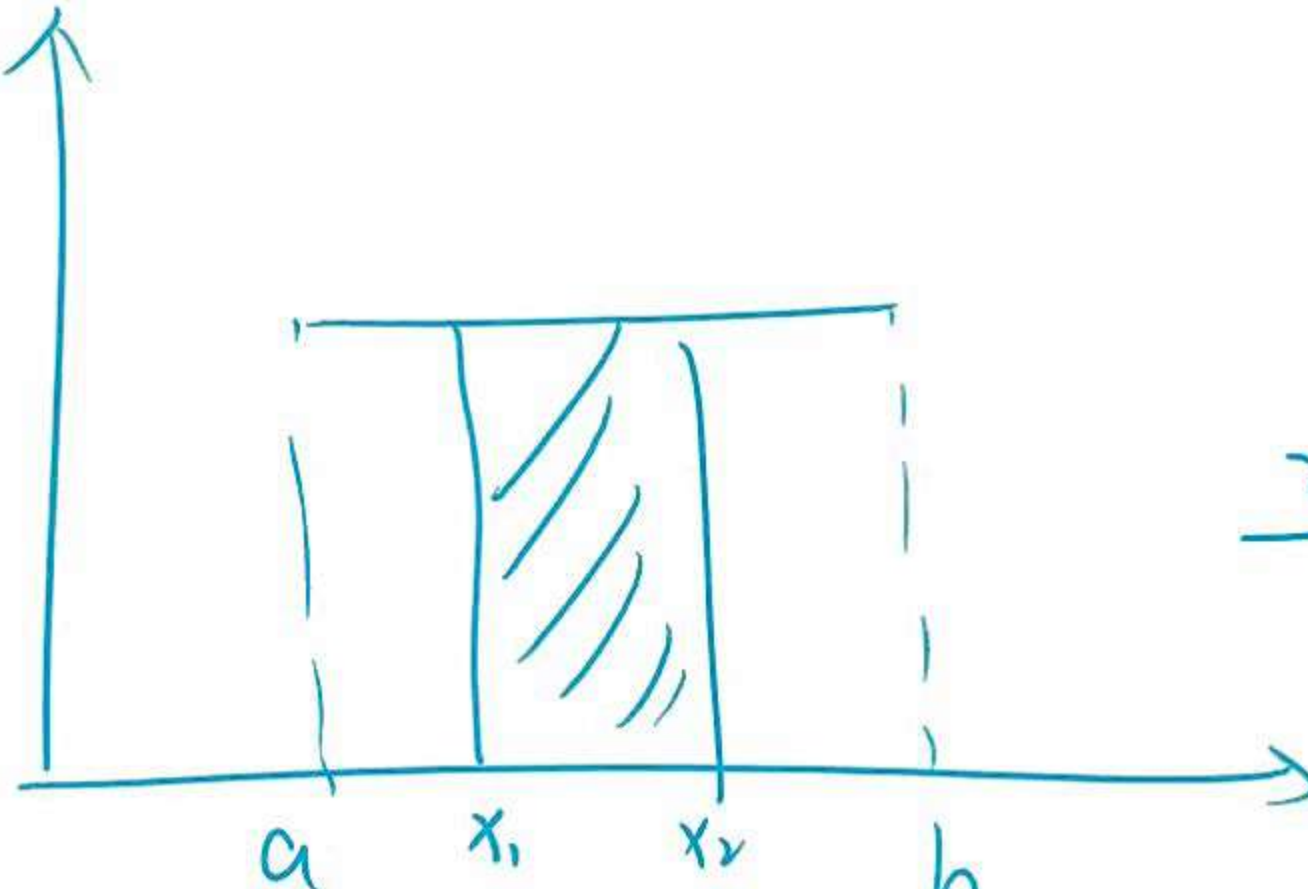
Continuous Uniform Distribution

- **Properties**
  - $E(X) = (a + b)/2$
  - $D(X) = (b - a)^2/12$
  - For all $a \le x_1 < x_2 \le b$, we have:

$$\boxed{P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx = \frac{x_2 - x_1}{b - a}}$$



- **Example**
  - eg: The random variable X with density function $f(x) = k/3$ for $2 \le x \le 8$, and 0 otherwise. Calculate its mean.



$$\frac{x_2 - x_1}{b - a}$$

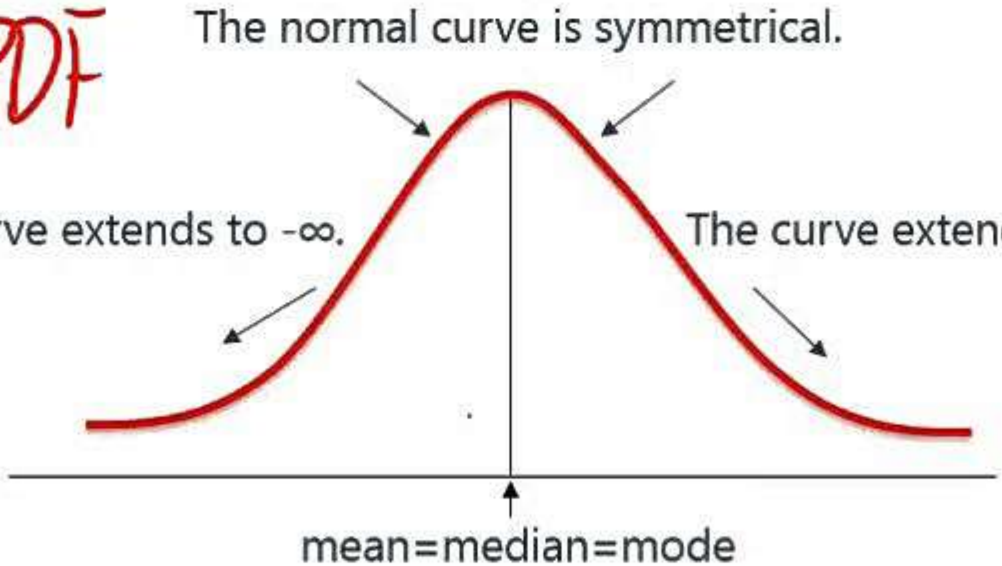
Normal Distribution

Normal Distribution

- As n increases, the binomial distribution approaches Normal Distribution.

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2\sigma^2}(x-u)^2}}$$

PDF



Normal Distribution

性质

- **Properties**
  - $X \sim N(\mu, \sigma^2)$, fully described by its two parameters μ and σ^2.
  - Bell-shaped, symmetrical distribution: skewness = 0; kurtosis = 3.
  - A linear combination (function) of two (or more) normally distribution random variables is itself normally distributed.
  - The tails get thin and go to zero but extend infinitely, asymptotic.

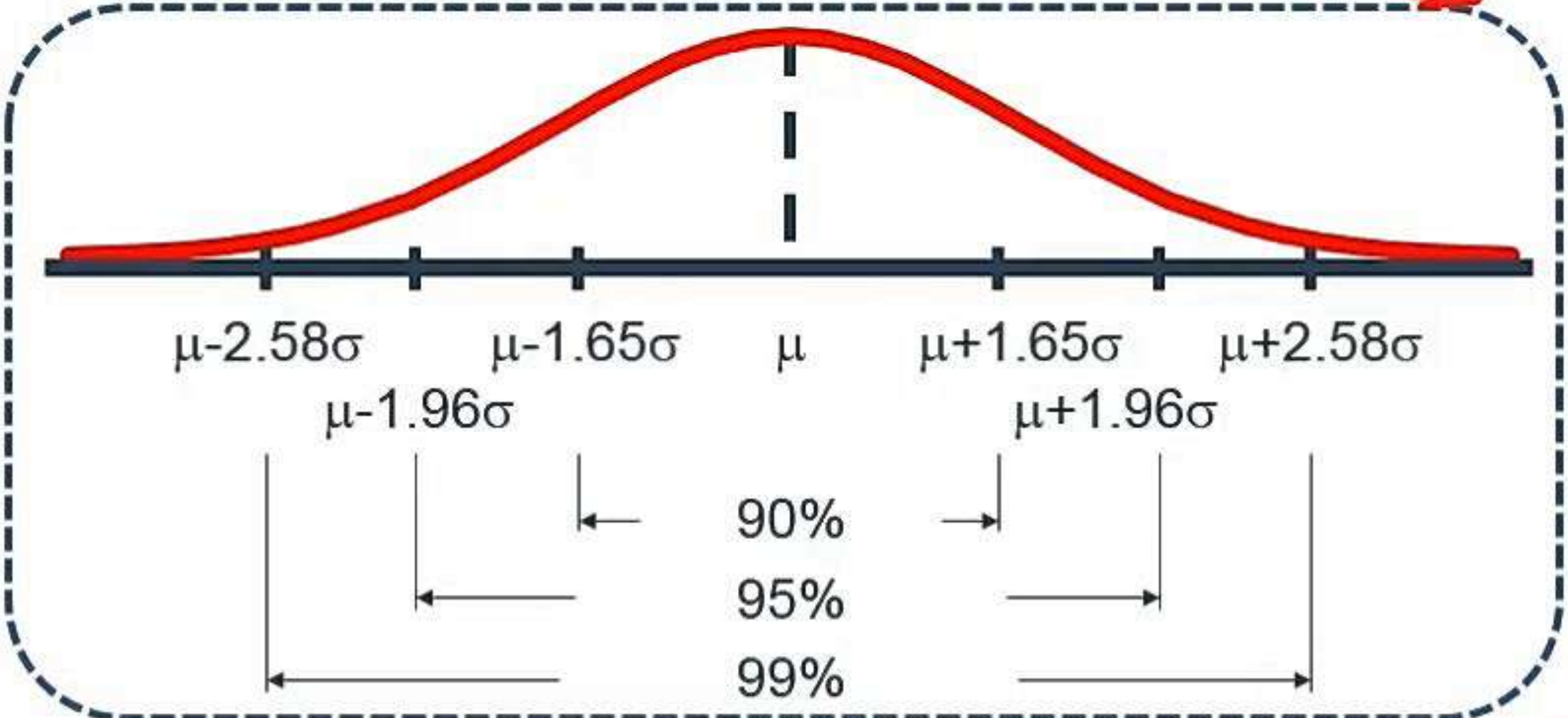
[illegible] $N(170, 2^2)$ [illegible] $N(1, 3^2)$

Normal Distribution

170

➢ **The confidence intervals** 置信区间 95% [165 , 175]

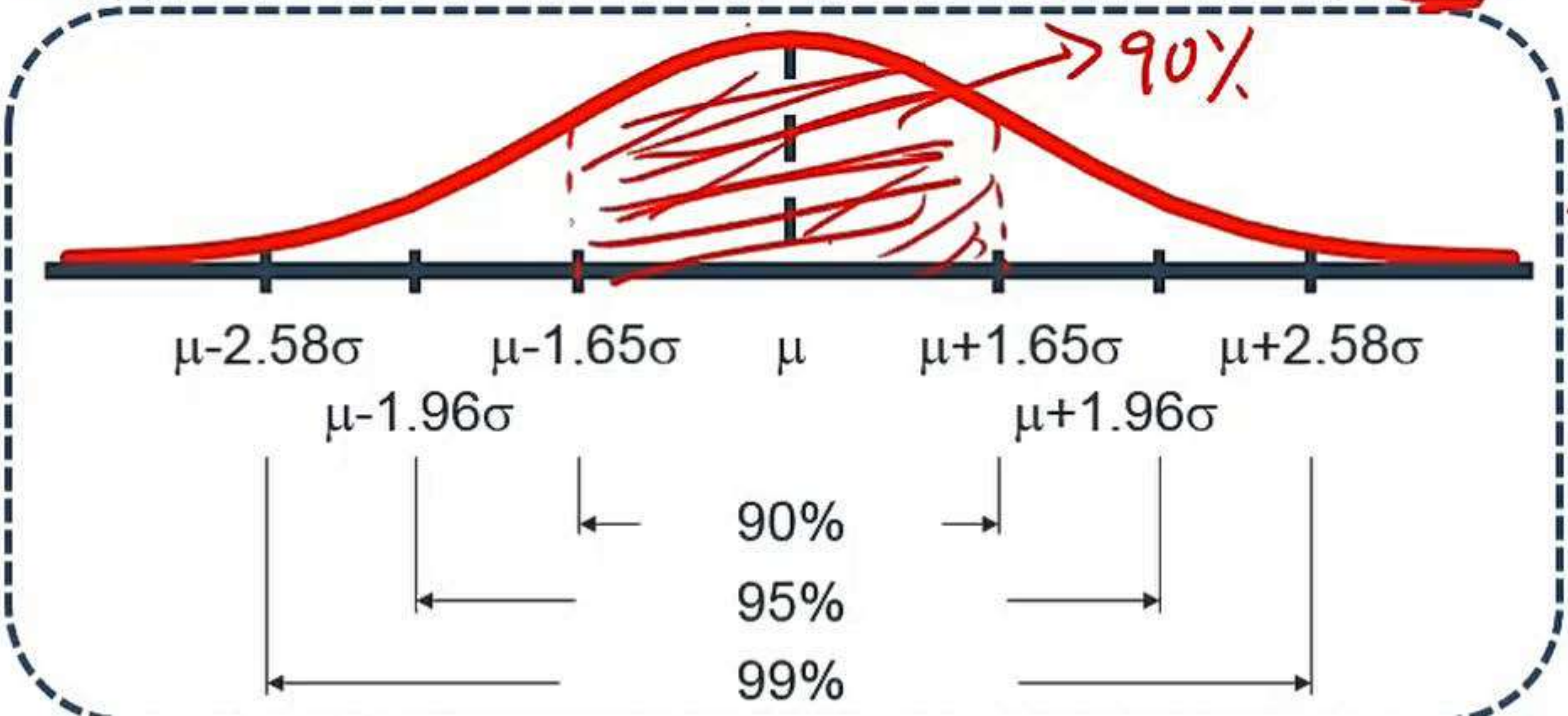
- Approximately 68% of all observations fall in the interval μ±σ
- Approximately 90% of all observations fall in the interval μ±1.65σ
- Approximately 95% of all observations fall in the interval μ±1.96σ
- Approximately 99% of all observations fall in the interval μ±2.58σ

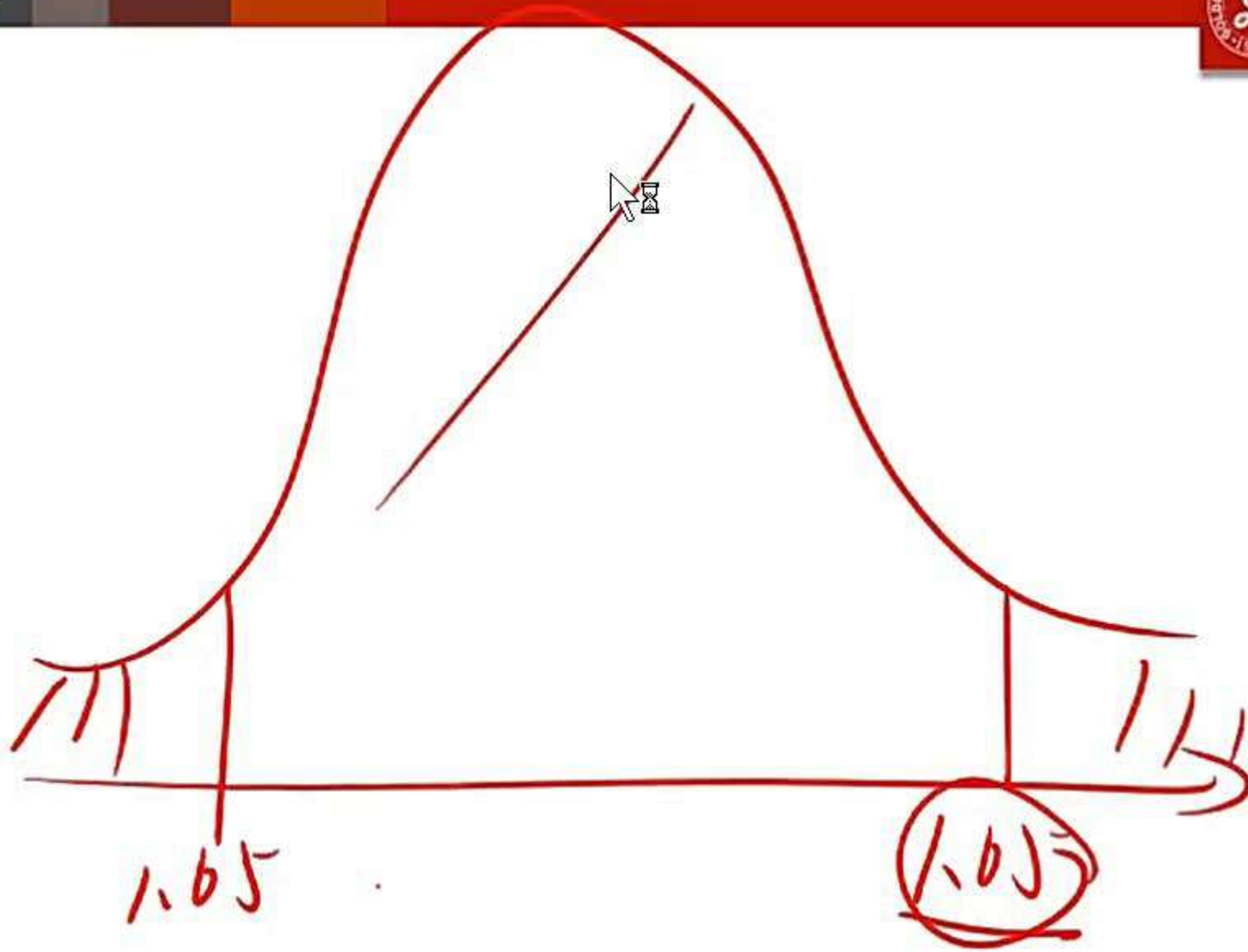


Normal Distribution

➢ **The confidence intervals** 置信区间 95% (165 , 175] 170

- Approximately 68% of all observations fall in the interval μ±σ
- Approximately 90% of all observations fall in the interval μ±1.65σ
- Approximately 95% of all observations fall in the interval μ±1.96σ 双尾
- Approximately 99% of all observations fall in the interval μ±2.58σ





1.65 → 右边 5%   2.33 → 右边 1%

1.96 → 右边 2.5%   2.58 → 右边 0.5%

The Standard Normal Distribution

- **The standard normal distribution**
  - N(0,1) or Z
  - Standardization: if X ~ N(μ, σ^2), then $\boxed{Z = \frac{X-\mu}{\sigma} \sim N(0,1)}$
  - Z-table
- **How we use the standard normal distribution to compute various probabilities?**
  - Example: X ~ N(70, 9), compute the probability of X ≤ 64.12.

    ✓ $Z = \frac{X-\mu}{\sigma} = \frac{64.12-70}{3} = -1.96$

    ✓ P(Z ≤ -1.96) = 0.0250
  - Question 1: compute the probability of X ≥ 75.9.
  - Question 2: compute the probability of 64.12 ≤ X ≤ 75.9.

A: $\mu = 170$ $\sigma = 5$

$\Downarrow$

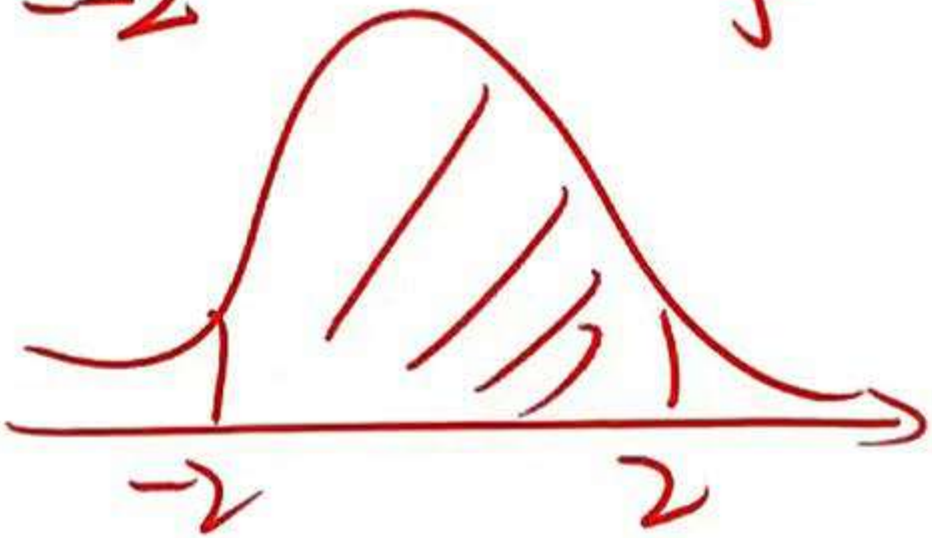
$\mu = 0$ $\sigma = 1$

B: $\mu = 168$ $\sigma = 6$

$[160 - 180]$

X X

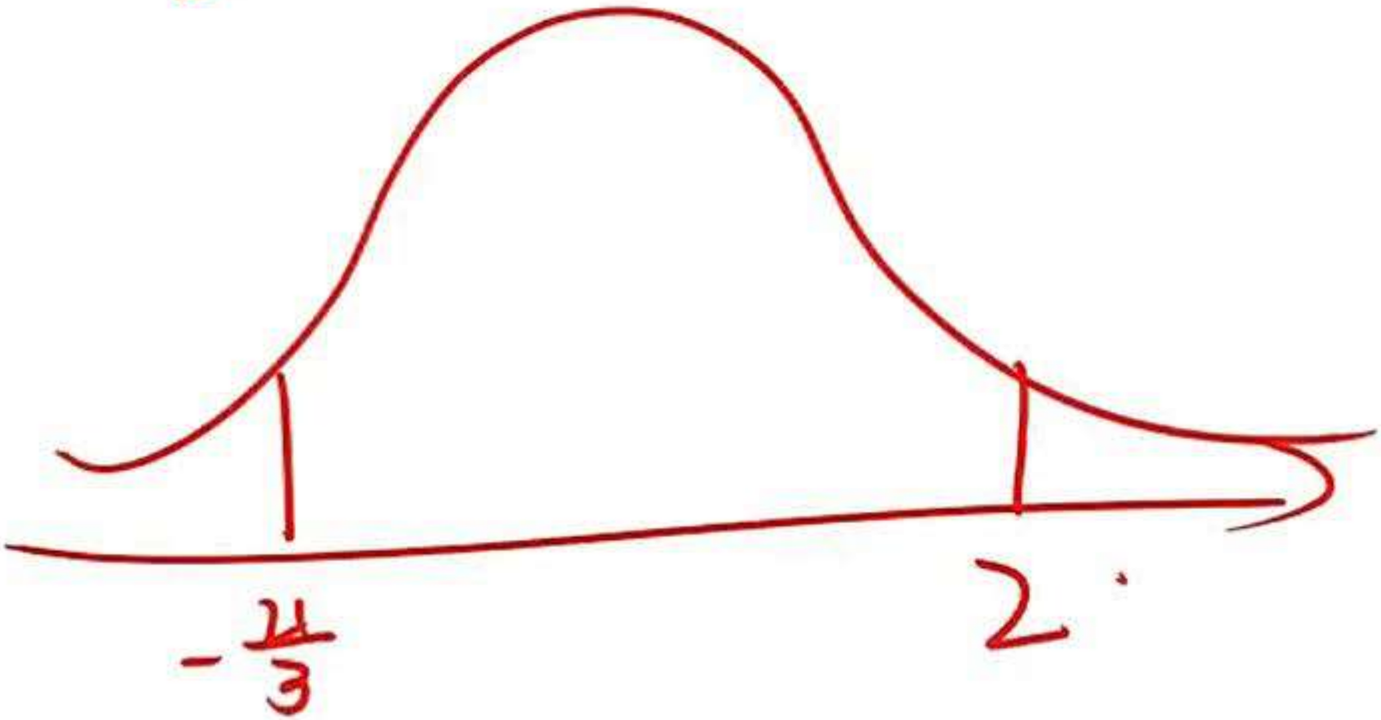
A: $\dfrac{160-170}{5} = -2$ $\dfrac{180-170}{5} = 2$



12

B. $\frac{160-168}{6} = -\frac{4}{3}$

$$\frac{180-168}{6} = 2.$$



The Standard Normal Distribution

160 – 180

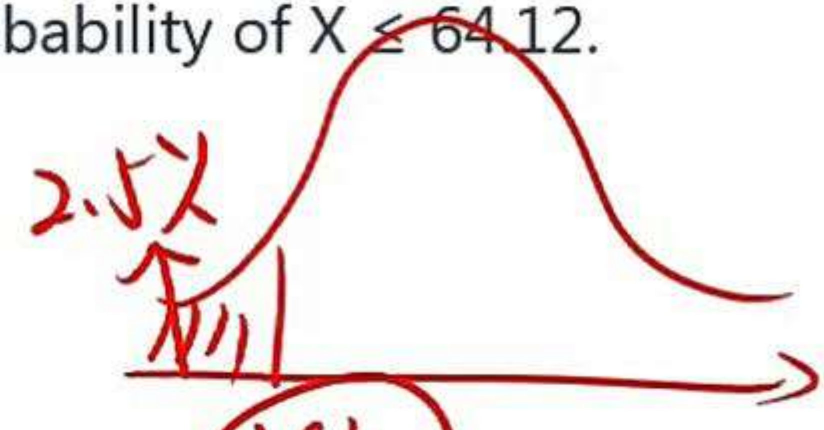
- **The standard normal distribution**
  - N(0,1) or Z
  - Standardization: if $X \sim N(\mu, \sigma^2)$, then $\boxed{Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)}$ ✓
  - Z-table

eg

- **How we use the standard normal distribution to compute various probabilities?**
  - Example: $X \sim N(70, 9)$, compute the probability of $X \le 64.12$.

    ✓ $Z = \dfrac{X-\mu}{\sigma} = \dfrac{64.12-70}{3} = -1.96$

    ✓ $P(Z \le -1.96) = 0.0250$



  - Question 1: compute the probability of $X \ge 75.9$. 1.96
  - Question 2: compute the probability of $64.12 < X < 75.9$.

75.9 – 70

Example 1

eg:

C

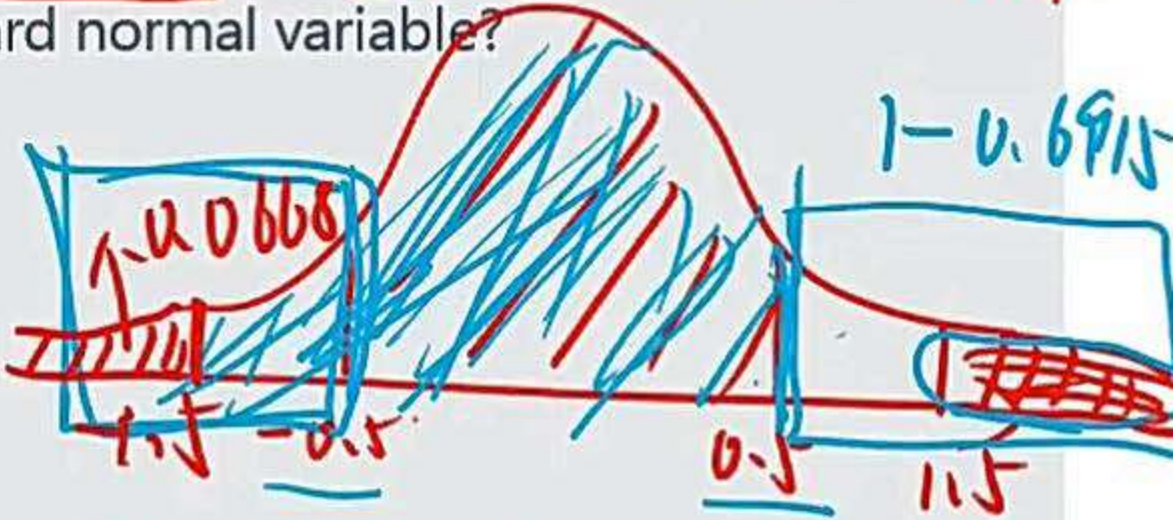
- Let Z be a standard normal random variable, and event X is defined to happen if either Z takes a value between -0.5 and +0.5 or Z takes any value greater then 1.5. What is the probability of event X happening if N(0.5) = 0.6915 and N(-1.5) = 0.0668, where N(.) is the cumulative distribution function of a standard normal variable? CDF

A. 0.2583

B. 0.3753

C. 0.4498

D. 0.7583



- Correct answer : C $0.6915 - [1 - 0.6915] + 0.0668$

Example 2

- Which type of distribution produces the lowest probability for a variable to exceed a special extreme value which is greater than the mean, assuming the distribution all have the same mean and variance?
  A. A leptokurtic distribution with a kurtosis of 4.
  B. A leptokurtic distribution with a kurtosis of 8.
  C. A normal distribution.
  D. A platykurtic distribution.

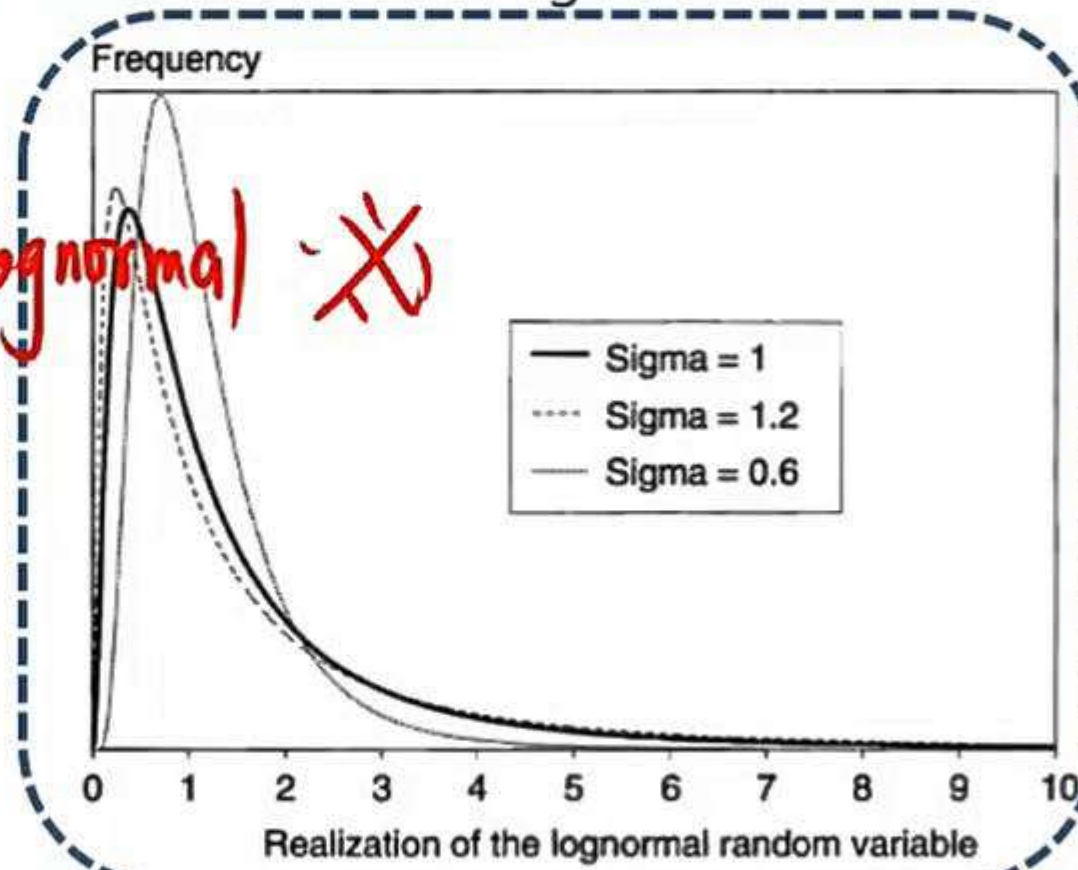
矮峰薄尾

- Correct Answer：D

五 Lognormal Distribution 对数正态分布

- **Lognormal Distribution**：描述股价变动的
  - The Black-Scholes Model assumes that the price of the underlying asset is lognormally distributed.
  - If lnX is normal, then X is lognormal; if a variable is lognormal, its natural log is normal. 定性 必
  - It is useful for modeling asset prices which never take negative values.
  - Right skewed.
  - Bounded from below by zero.

$\ln X \sim$ normal $\Longleftrightarrow X \sim$ lognormal 必



$\ln X \sim$ normal $\longleftrightarrow$ $X \rightarrow$ lognormal

eg (错)

$\ln X \sim$ lognormal $\longrightarrow$ $X \sim$ normal

$X > 0$

Chi-Square Distribution

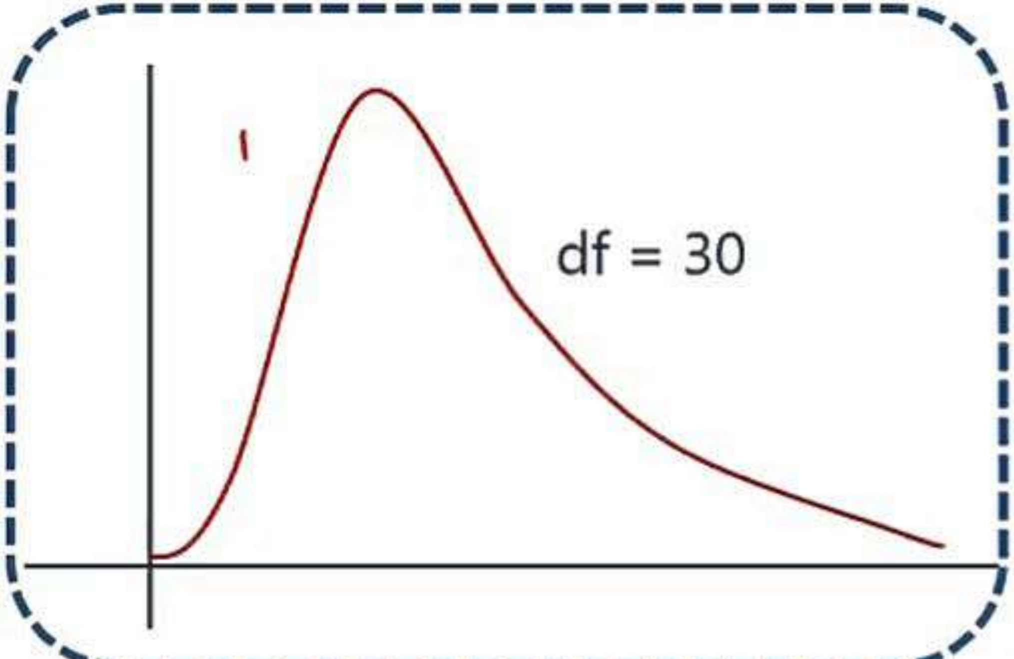
- **Chi-Square (χ^2) Distribution** :非负数

  - Chi-Square test statistic, χ^2, with n – 1 degrees of freedom, is computed as:

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

  - Notice

$$\sum Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2 \sim \chi^2_{(k)}$$



七 *t* Distribution：与正态分布极度相似.

$N(\mu, \sigma^2)$

- **t Distribution (student's t distribution)**

  - Recall that, $Z = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)$, both μ_X and $\sigma_X{}^2$ are known.

  - Suppose we only know μ_X and estimate $\sigma_X{}^2$ by its (sample) estimator $S_X^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$, we obtain a new variable.

$$t = \frac{\bar{X} - m_X}{S_X/\sqrt{n}} \sim t_{n-1}$$

- **Explain the d.f. (degrees of freedom)**

  - Before we compute the $S_X{}^2$ (and hence S_X) , we must first compute $\bar{X}$. But since we use the same sample to compute $\bar{X}$, we have (n-1), not n, independent observations to compute $S_X{}^2$, so to speak, we lose 1d.f.

t Distribution：与正态分布相度相似.

$N(\mu, \sigma^2)$

- **t Distribution (student's t distribution)**

z-statistics 统计量

  - Recall that, $\boxed{Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \sim N(0,1)}$, both μ_X and ${\sigma_X}^2$ are known.

  - Suppose we only know μ_X and estimate ${\sigma_X}^2$ by its (sample) estimator $S_X^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$, we obtain a new variable.

t-statistics

$$t = \frac{\bar{X} - m_X}{S_X / \sqrt{n}} \sim t_{n-1}$$

- **Explain the d.f. (degrees of freedom)**

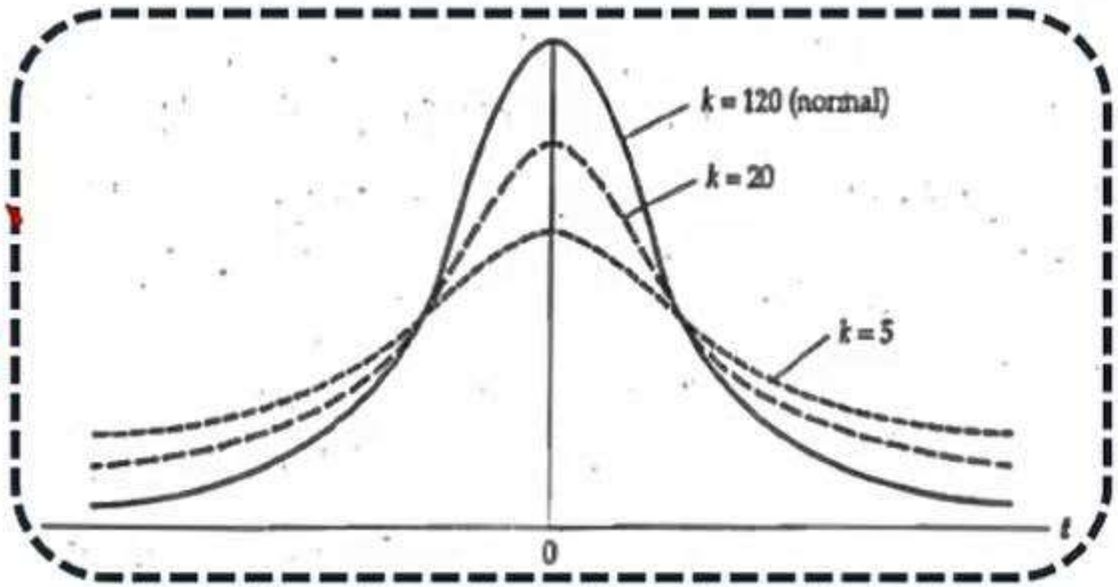
  - Before we compute the ${S_X}^2$ (and hence S_X) , we must first compute $\bar{X}$. But since we use the same sample to compute $\bar{X}$, we have (n-1), not n, independent observations to compute ${S_X}^2$, so to speak, we lose 1d.f.

t Distribution

- **Property**
  - Symmetric.
  - The mean of t distribution is zero, and its variance n/(n − 2).
  - The variance of t distribution is larger than the variance of the standard normal distribution, so t distribution is flatter than the normal distribution, but as n increases, the variance of t distribution approaches the variance of the standard normal distribution, namely 1.

① 矮峰肥尾

② $n \uparrow \sim N$



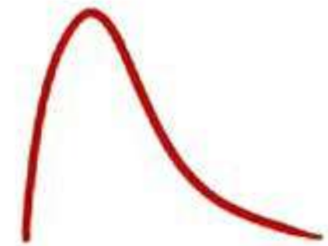
F-Distribution

z t

χ^2 F

- **F-Distribution**
  - If U_1 and U_2 are two independent Chi-Squared distributions with k_1 and k_2 degrees of freedom, respectively, then X:

$$\boxed{X = \frac{U_1/k_1}{U_2/k_2} \sim F(k_1, k_2)}$$



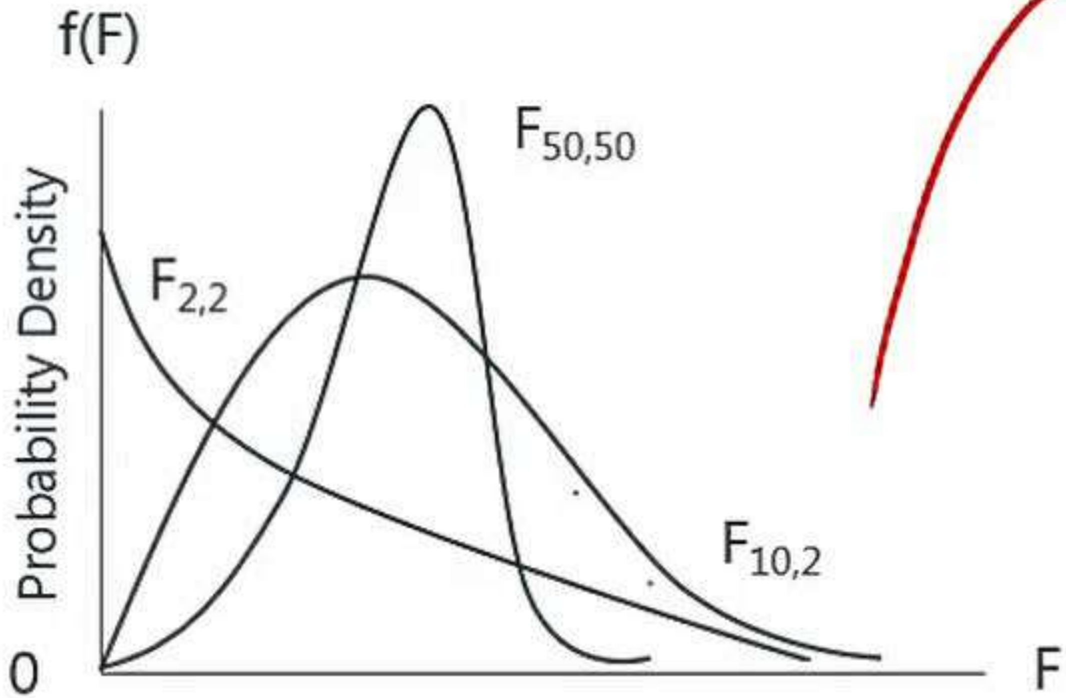
follows an F-distribution with parameters k_1 and k_2.

  - As d.f. increase, the F-Distribution approaches Normal Distribution.
  - If X is a random variable with a t distribution with k degrees of freedom, then X_2 has an F-Distribution with 1 and k degrees of freedom:

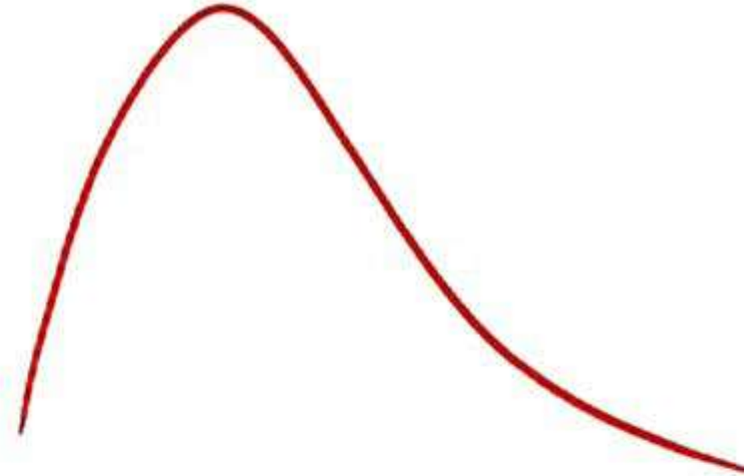
$$\boxed{X^2 \sim F(1, k)}$$

F-Distribution

- **F-Distribution**



The F distribution for various d.f.



- **Properties**
  - Skewed to the right and also ranges between 0 and infinity.
  - Approaches the Normal Distribution as k_1 and k_2, the d.f. become large.

Example 1

- On a multiple choice exam with four choices for each of six questions, what is the probability that a student gets less than two questions correct simply by guessing?
  - A. 0.46%
  - B. 23.73%
  - C. 35.60%
  - D. 53.39%

0 , 1

$C_6^0 (25\%)^0 (75\%)^6$

$C_6^1 (25\%)^1 (75\%)$

- Correct Answer : D
  - $p(X=0) = (3/4)^6 = 17.80\%$
  - $p(X=1) = 6 \times (1/4) \times (3/4)^5 = 35.59\%$
  - The probability of getting less than two questions correct is $p(X=0) + p(X=1) = 53.39\%$.

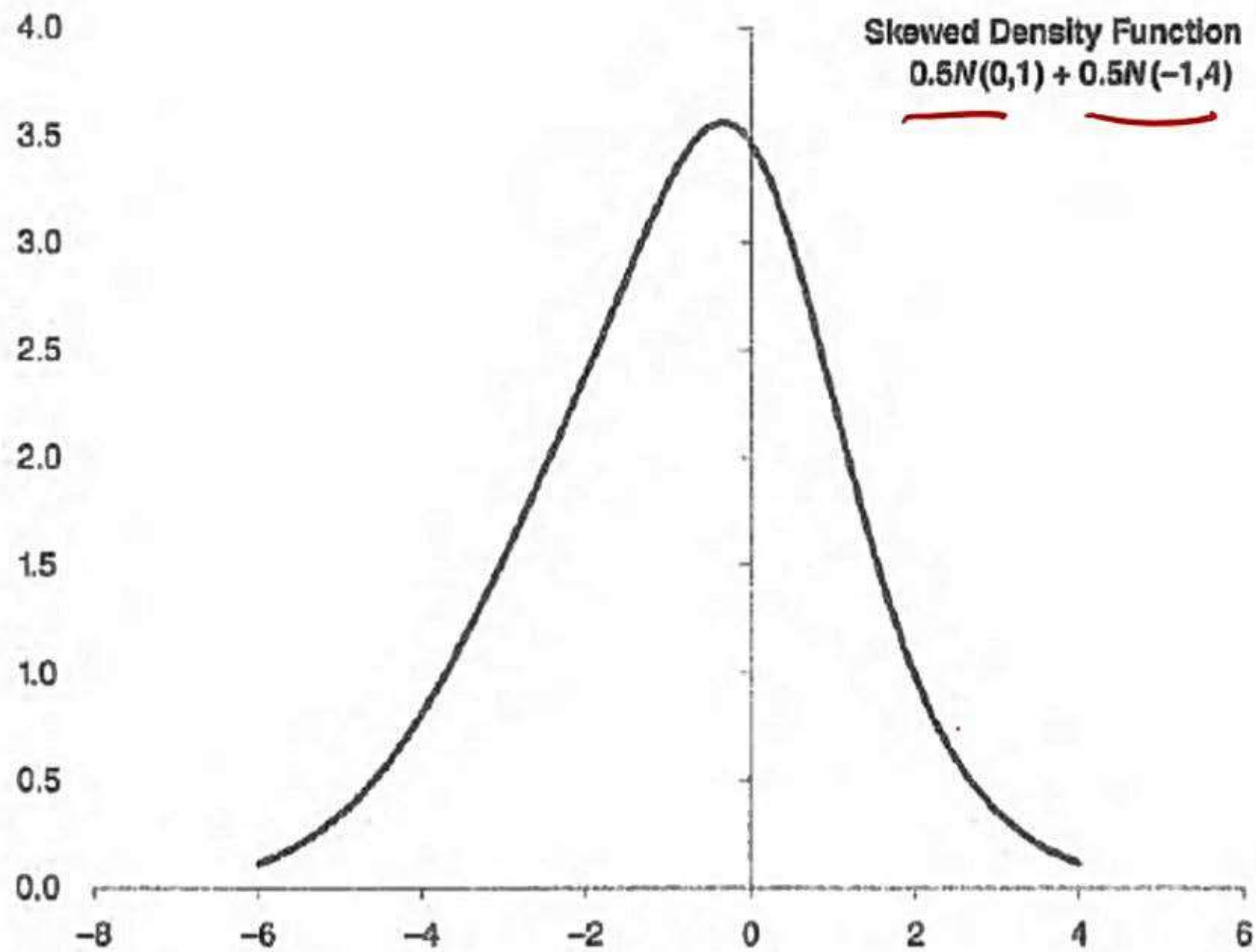
Mixture distribution

- The distribution that results from a weighted average distribution of density functions is known as a mixture distribution. More generally, we can create a distribution:

$$f(x) = \sum_{i=1}^{n} wi\, fi(x) \quad s.t. \sum_{i=1}^{n} w\, i = 1$$

  - where the various $f_i(x)$'s are known as the component distributions, and the w_i's are known as the mixing proportions or weights.

Mixture distribution



GOLDEN FUTURE

总结：

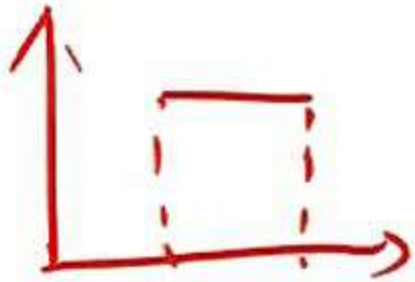
1. binormial $\begin{cases} \text{mean} = np \\ \text{ariance} = np(1-p) \\ 计算：C_n^k\ p^k (1-p)^{n-k} \end{cases}$

2. Possion $\begin{cases} \left.\begin{aligned} \text{mean} &= \lambda \\ \text{Variance} &= \lambda \end{aligned}\right\} (p \to 0,\ n \to \infty,\ np = \text{constant}) \\ 计算\ P(X=k) = \dfrac{\lambda^k}{k!} e^{-\lambda}. \end{cases}$

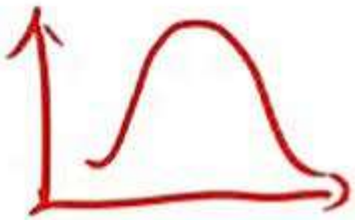
离散的

3. continuous uniform

- mean = $\frac{a+b}{2}$.
- 图



4 normal

- 图: 
- 性质:
- 标准化: $\frac{X-\mu}{\sigma}$

5 lognormal

- 用途: 描述 stock price.
- 性质:
  - 取值范围
  - 右偏.
  - $\ln X \longleftrightarrow X$

6 Chi-square
- 定义:
- 特点
  - 取值
  - 右偏
  - $n\uparrow \rightarrow$ normal).
- 计算

7. t-distribution
- 和 normal 几乎一样.
- (当 σ^2 未知 用 s^2 来代替)
- mean $= 0$
- variance $= \dfrac{n}{n-2}$

8 F-distribution
- 图 
- 特点
  - 取值范围
  - $n\uparrow \rightarrow$ normal.

①总体不可测
② 检验是错误性的

六、Type I & Type II error

Hypothesis Tests and Confidence Intervals

一、Term

二、CLT: centeral limit thereom

三、point estimate & confidencen interval estimate

四 Steps: 五步、 五 p-value

Sample and Population

- **Sampling and Estimation**
  - 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).

描述性统计学、

$$
\begin{tikzcd}
\boxed{\text{population}} \arrow[d, "\text{sampling}"'] & \boxed{\text{population parameter}} \\
\boxed{\text{sample}} & \boxed{\text{sample statistic}} \arrow[u, "\text{estimation}"']
\end{tikzcd}
$$

抽样

估计

总体参数：描述总体数据特征

样本统计量：描述样本数据特征

推断性统计学

Statistical Inference: Estimation and Hypothesis Testing

- **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** 样本误
  - sampling error of the mean = sample mean – population mean
- **The sample statistic itself is a random variable and has a probability distribution.** 样本统计量本身就是随机变量

$\bar{X}_1 = 170$ prob

$\bar{X}_2 = ?$

$\bar{X}$

The Central Limit Theorem

The Central Limit Theorem (CLT) 1.作用：将未知分布变成正态

- If X_1, X_2, … ,X_n a random sample from any population (i.e.,分布 probability distribution) with mean μ_x and ${\sigma_x}^2$, the sample mean $\bar{X}$ tends to be normally distributed with mean μ_x and $\frac{{\sigma_x}^2}{n}$ variance, as the sample size increases indefinitely (technically, infinitely) (≥ 30).
- Of course, if the X_i happen to be from the normal population, the sample mean follows the normal distribution regardless of the sample size.
- Standard Error (SE) of mean $\bar{X}$: $\mathrm{SE}(X) = \frac{s}{\sqrt{n}}$
  - ✓ However, the population's standard deviation is almost never known. Instead, we use the standard deviation of the sample mean.

CLT：

2.定义：

未知分布 $X_1 \quad X_2 \cdots X_n \qquad \mu \qquad \sigma^2$

抽 $n=50 \quad \overline{X_1} = a$

$n=50 \quad \overline{X_2} = b$

$\vdots \qquad \vdots$

$\overline{X_n}$

$$\boxed{\overline{X_i} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$

3. 性质

① $E(\bar{x}_i) = \mu$

② $\sigma^2_{\bar{x}_i} = \frac{\sigma^2}{n}$ √

③ $n \geq 30$

样本容量

x_i

$\sigma_{\bar{x}_i}$: standard error

标准误

Point Estimation and Confidence Interval Estimate

➢ Point Estimation 点估计

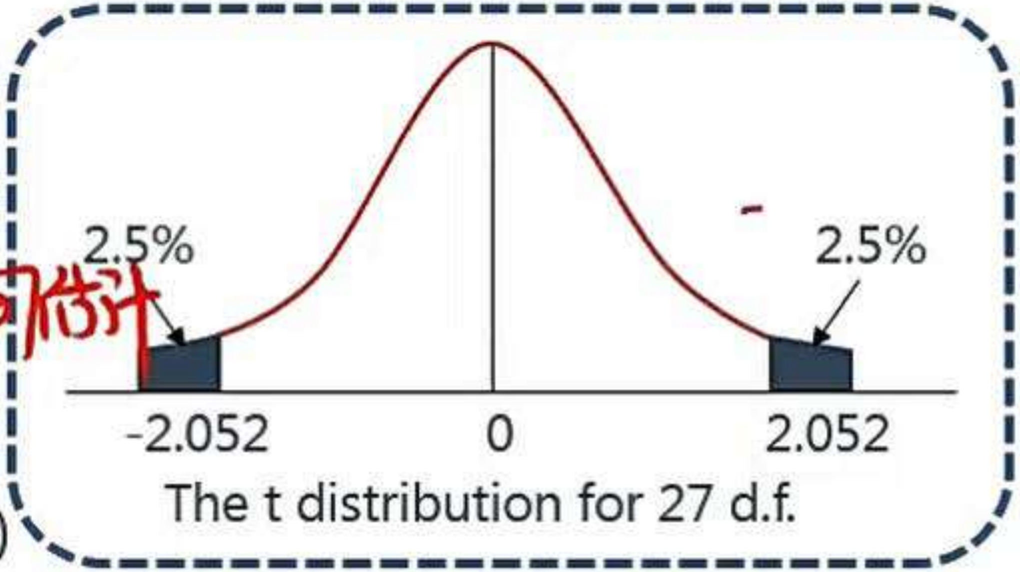
- Using a single numerical value to estimate the parameter of the population.

parameter

$$\boxed{\bar{X} \rightarrow \mu_X \qquad\qquad S^2 \rightarrow \sigma_X^2}$$

Estimator

➢ Confidence Interval Estimate 区间估计



- Level of significance (alpha)
- Degree of confidence (1 – alpha)
- Confidence Interval = [Point Estimate +/- (reliability factor) × standard error]

查表

$\bar{X} \pm k \cdot SE$ — standard error 标准误

置信因子

critical vale 关键值

分位点

1.65

1.96

2.58

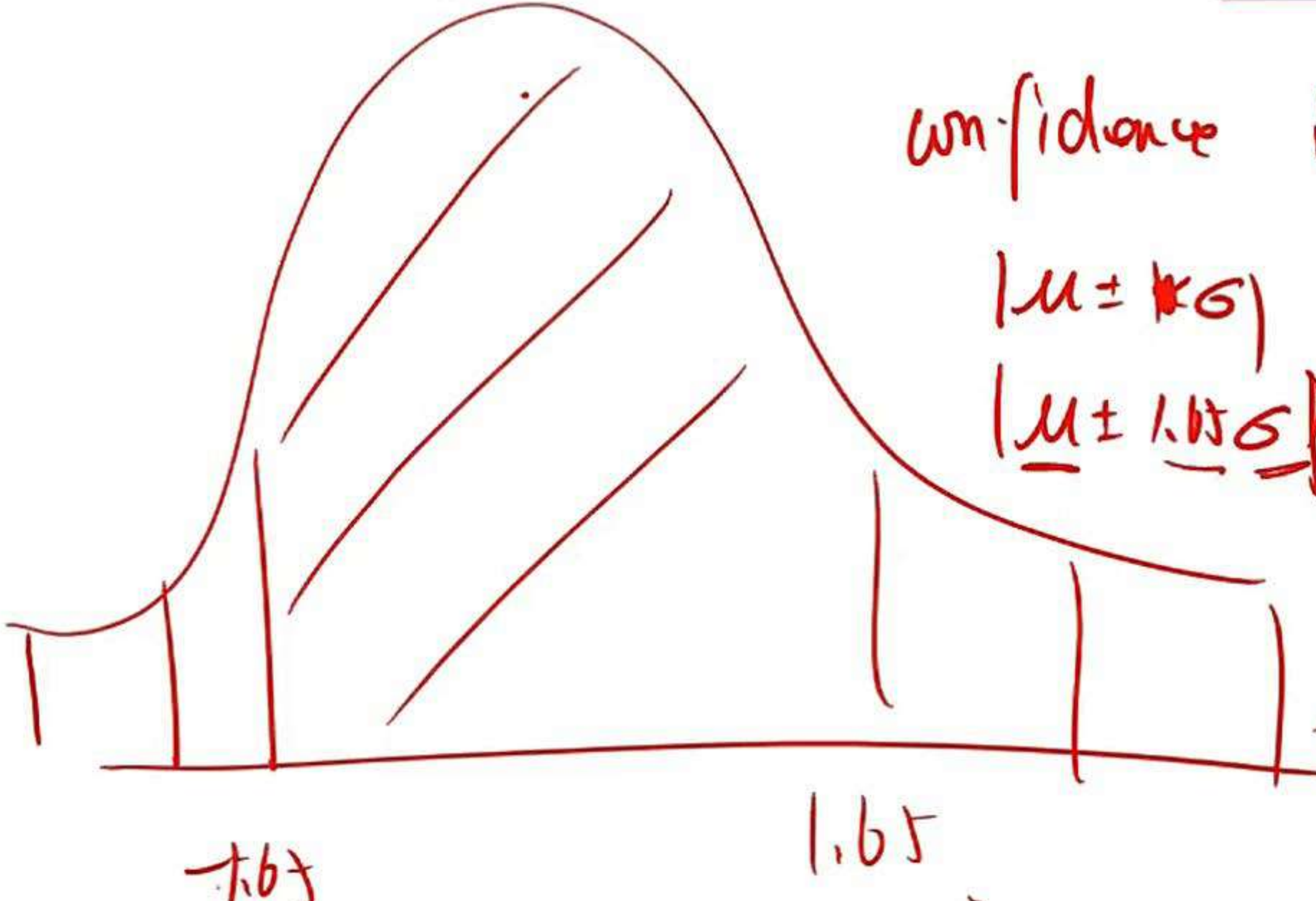
2.33

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

confidence interval

$[\mu \pm$ [illegible] $\sigma]$

$[\mu \pm 1.65\sigma]$



Confidence Intervals

$\bar{X} \pm k \cdot SE$

- **Confidence Interval Estimation**
- **The population has a normal distribution with a known variance.**
  - Confidence interval:

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

Point estimate Reliability factor Standard error

- **The population has a normal distribution with a unknown variance.**
  - Confidence interval:

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

When sampling form a:	small sample (n < 30)	larger sample (n ≥ 30)
Normal distribution with known variance	z-Statistic	z-Statistic
Normal distribution with unknown variance	t-Statistic	t-Statistic or z-Statistic
Nonnormal distribution with known variance	not available	z-Statistic
Nonnormal distribution with unknown variance	not available	t-Statistic or z-Statistic

GOLDEN FUTURE

Confidence Intervals

$\bar{x} \pm k \cdot SE$

- **Confidence Interval Estimation**
- **The population has a normal distribution with a known variance.**
  - Confidence interval:

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

Point estimate   Reliability factor   Standard error

- **The population has a normal distribution with a unknown variance.**
  - Confidence interval:

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

When sampling form a:	small sample (n < 30)	larger sample (n ≥ 30)
Normal distribution with known variance	z-Statistic	z-Statistic
Normal distribution with unknown variance	t-Statistic	t-Statistic or z-Statistic
Nonnormal distribution with known variance	not available	z-Statistic
Nonnormal distribution with unknown variance	not available	t-Statistic or z-Statistic

① $n \geq 30 \rightarrow Z$

② σ^2已知 Z，σ^2未知 t

③ 非正态小样本不可估.

Statistical Inference: Estimation and Hypothesis Testing

- Estimation and Hypothesis Testing: Twin Branches Of Statistical Inference

Normal

variance 未知

45人

Price to earning (P/E) ratios of 28 companies on the New York stock exchange (NYSE)			
Company	P/E	Company	P/E
AA	27.96	INTC	36.02
AXP	22.90	IBM	22.94
T	8.30	JPM	12.10
BA	49.78	JNJ	22.43
CAT	24.88	MCD	22.13
C	14.55	MRK	16.48
KO	26.22	MSFT	33.75
DD	28.21	MMM	26.05
EK	34.71	MO	12.21
XOM	12.99	PG	24.49
GE	21.89	SBC	14.87
GM	9.86	UTX	14.87
HD	20.26	WMT	27.84
HON	23.36	DIS	37.10
Mean = 23.25	Variance = 90.13	Standard deviation = 9.49	

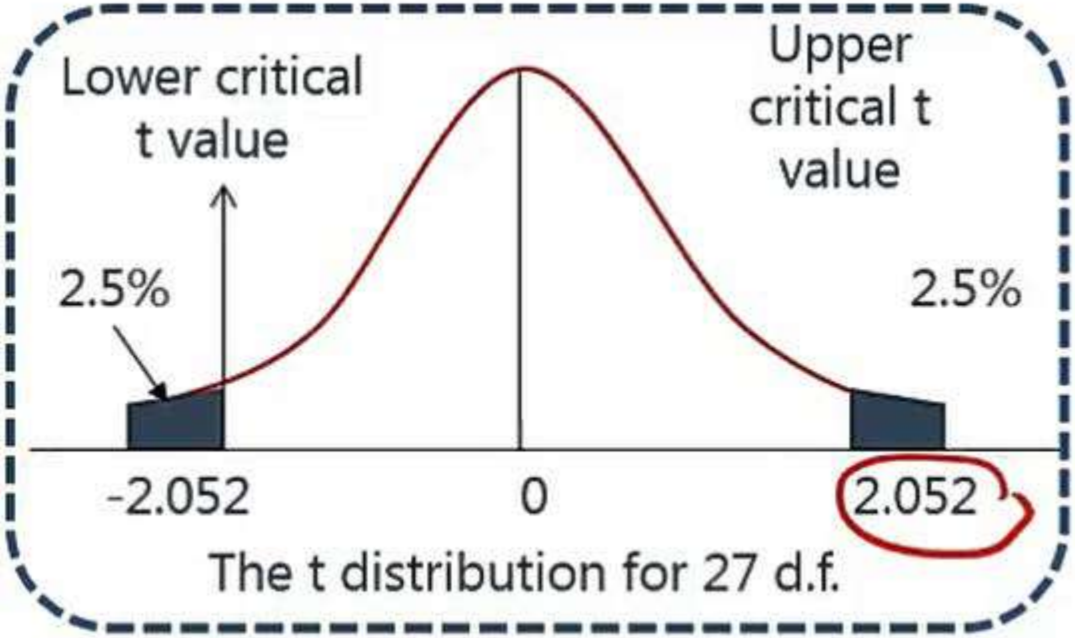
Statistical Inference: Estimation and Hypothesis Testing

- **Example: Confidence Interval Estimation**
  - Statistic：统计量 σ

$$t = \frac{\bar{X} - \mu_x}{S_x/\sqrt{n}} \sim t(n-1)$$

  - Critical Value

$$P(-2.052 \le t \le 2.052) = 0.95$$



  - Obtain a Random Interval

$$P\left(\bar{X} - 2.052\frac{S_X}{\sqrt{n}} \le \mu_X \le \bar{X} + 2.052\frac{S_X}{\sqrt{n}}\right) = 0.95$$

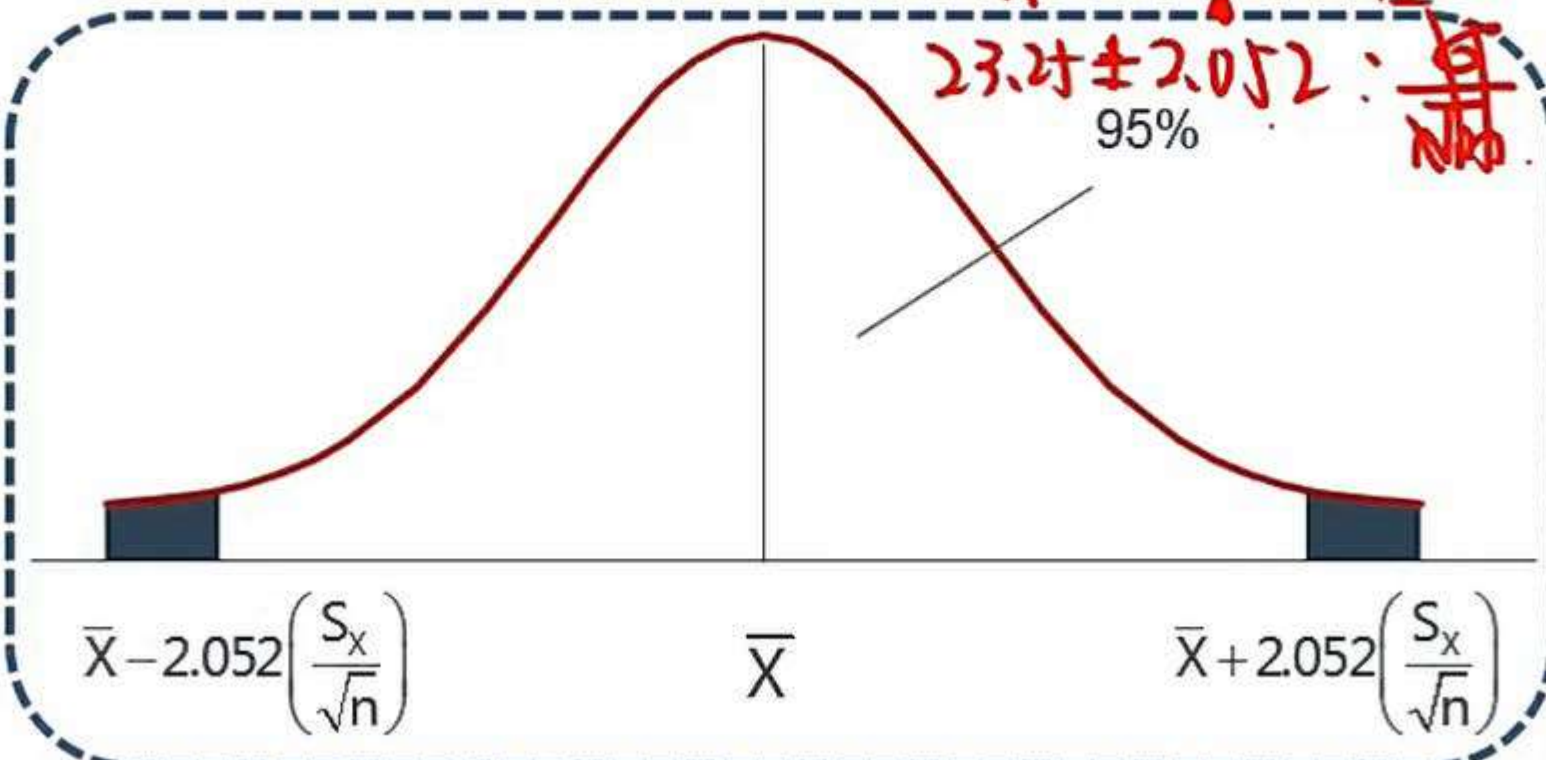
- **Important point: one should say that the probability is 0.95 that random interval contains the true μ_X.**

Statistical Inference: Estimation and Hypothesis Testing

- Using sample estimator, obtaining a confidence interval.
  - ✓ Returning to our P/E example, we have n = 28, $\bar{X} = 23.25$, and $S_X = 9.49$. We obtain $\boxed{19.57 \le \mu_X \le 26.93}$ as the 95% confidence interval for μ_X.

$\bar{X} \pm K \cdot SE$

$23.25 \pm 2.052 \cdot \frac{S}{\sqrt{n}} \cdot \frac{9.49}{\sqrt{28}}$



Hypothesis Testing

$\bar{X} = 23.25$

Step 1: State null and alternative hypotheses

原假设：希望拒绝的放入原假设

备择假设：把想要证明的放入备择假设

Step 2: Identify the test statistic

Step 3: Select a level of significance

"="

Step 4: Formulate a decision rule

Step 5: Take a sample, arrive at decision → Do not reject / Reject

less than or equal to

alter

Hypothesis Testing

$\bar{X} = 23.25$ ∞ K

Step 1

State null and alternative hypotheses

原假设：希望拒绝的放入原假设

备择假设：把想要证明的放入备择假设

Step 2

Identify the test statistic

$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

Step 3

Select a level of significance

显著性

Step 4

Formulate a decision rule

K

Step 5

Take a sample, arrive at decision

- Do not reject
- Reject

less than or equal to

alter

Hypothesis Testing

- **The null hypothesis (H_0) and alternative hypothesis (H_a)**
- **One-tailed test vs. Two-tailed test**
  - One-tailed test

    $H_0: \mu \geq 0 \qquad H_a: \mu < 0$

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

  - Two-tailed test

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

- **Critical Value**
  - The distribution of test statistic (z, t, χ^2, F)
  - Significance level (α)
  - One-tailed or two-tailed test

Hypothesis Testing

H_1

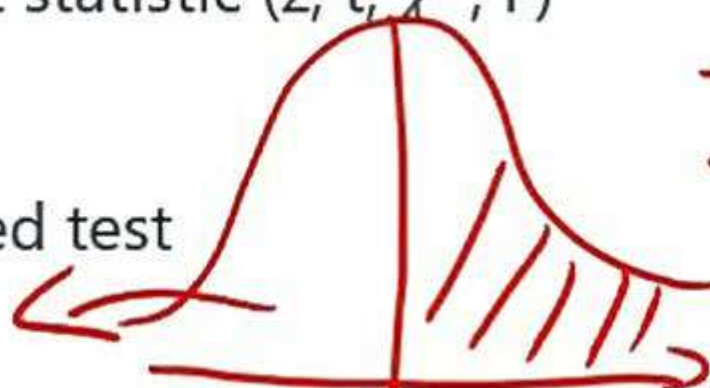
- **The null hypothesis (H_0) and alternative hypothesis (H_a)**
- **One-tailed test vs. Two-tailed test**
  - One-tailed test ：拒绝域只有1个　　$\bar{x} = 180$

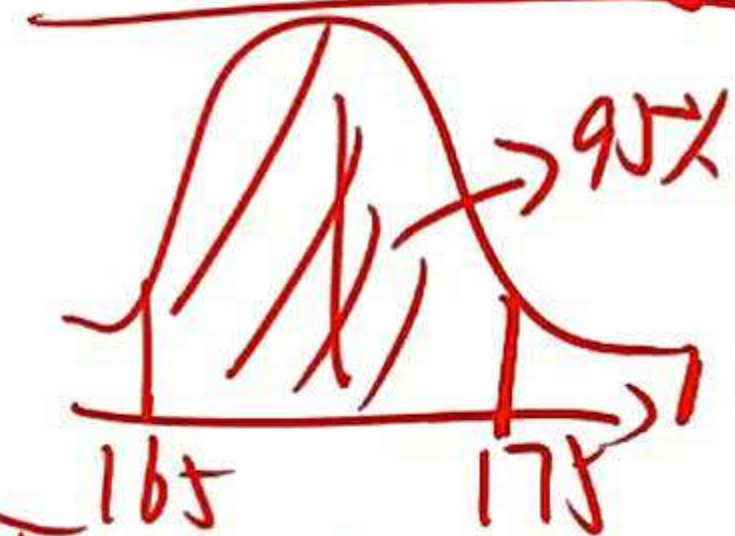
H_0: μ ≥ 0 + H_a: μ < 0		H_0: μ ≤ 0　　H_a: μ > 0

  - Two-tailed test ：拒绝域有2个　95%　165 − 175

H_0: μ = 0　　H_a: μ ≠ 0

- **Critical Value**
  - The distribution of test statistic (z, t, χ^2, F)
  - Significance level (α)
  - One-tailed or two-tailed test





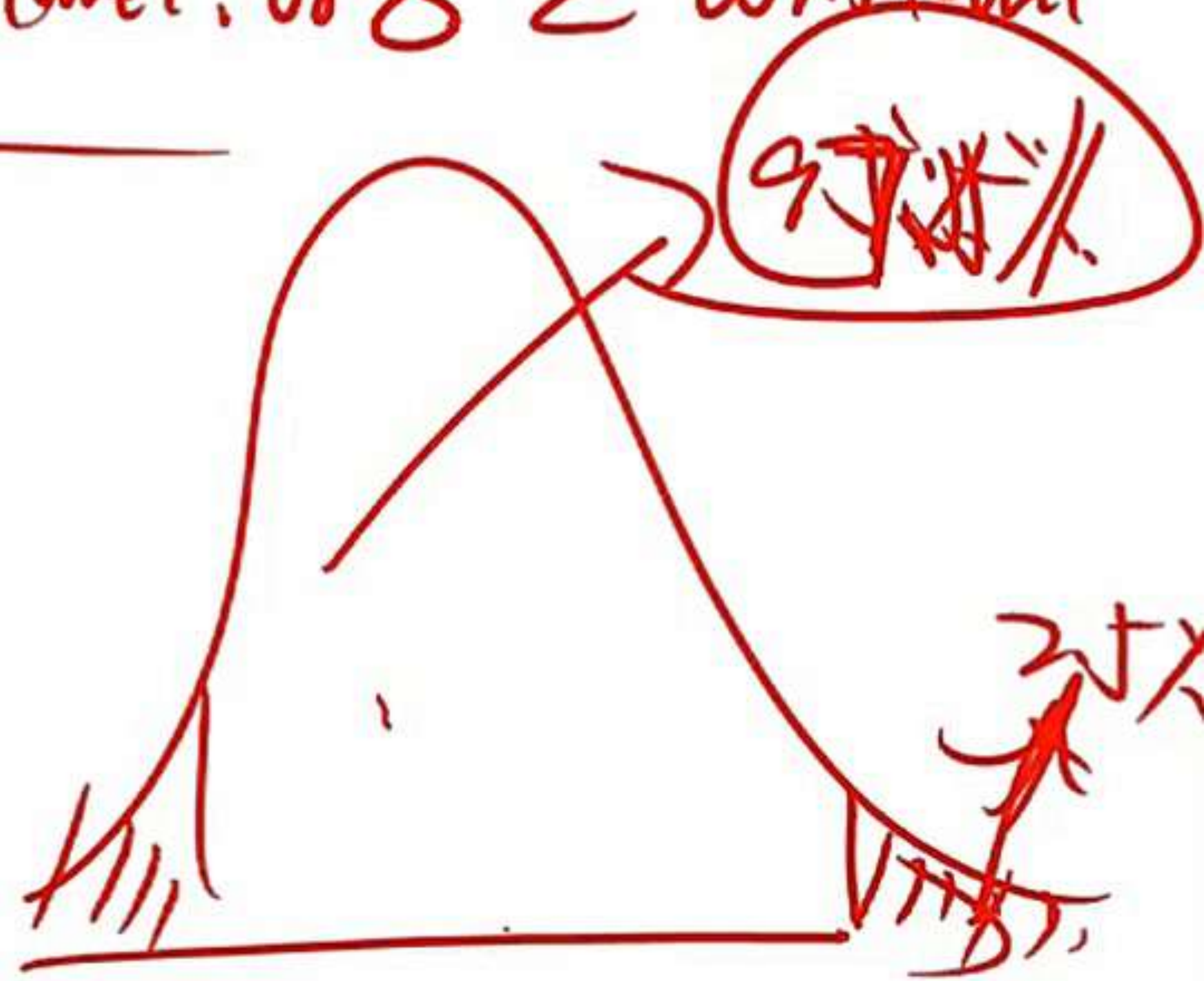
F

H_0: $\sigma = \text{constant}$

H_1: $\sigma > \text{constant}$, or $\sigma < \text{constant}$

$$F = \frac{S_1^2}{S_2^2}$$

F是单尾test



Hypothesis Testing

H_1

- The null hypothesis (H_0) and alternative hypothesis (H_a)
- One-tailed test vs. Two-tailed test
  - One-tailed test：拒绝域只有1个　　$\bar{X} = 180$

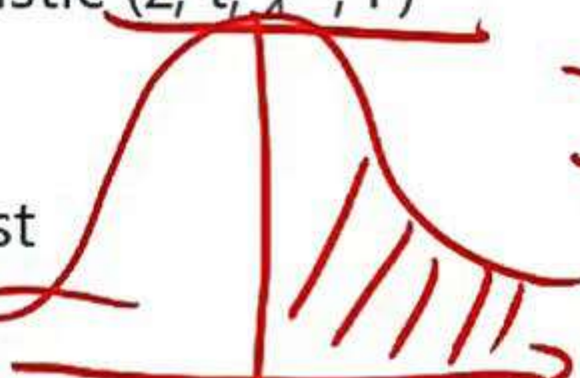
$H_0: \mu \geq 0$ + $H_a: \mu < 0$	$H_0: \mu \leq 0$　　$H_a: \mu > 0$

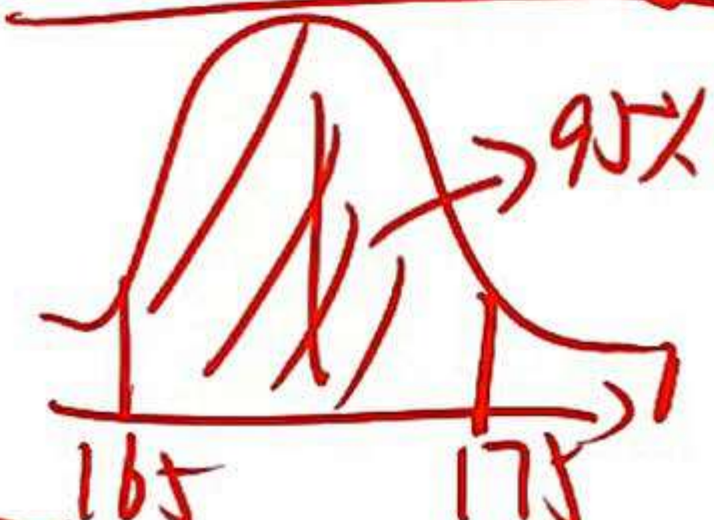
  - Two-tailed test：拒绝域有2个　95% [165 − 175]

$H_0: \mu = 0$　　$H_a: \mu \neq 0$

- Critical Value

  X　σ^2
  - The distribution of test statistic (z, t, χ^2, F)
  - Significance level (α)
  - One-tailed or two-tailed test





Statistical Inference: Estimation and Hypothesis Testing

➢ The Test of Significance Approach to Hypothesis Testing

- Steps:

$\mu = 18.5$

① State the null hypothesis H_0 and alternative hypothesis H_1

$$H_0 : \mu_X = 18.5, \quad H_1 : \mu_X \neq 18.5$$

② Select the test statistic and determine the distribution

Test statistic = (sample statistic – hypothesized value)/(standard error of the sample statistic)

$$t = \frac{(\bar{X} - \mu_X)}{S_X/\sqrt{n}} \sim t_{(n-1)} \checkmark$$

$$\frac{23.25 - 18.5}{9.49/\sqrt{38}} = 2.65$$

③ Choose the level of significance α(5%)

④ Obtain critical t value $t_{\frac{\alpha}{2}(n-1)} = 2.052$

Statistical Inference: Estimation and Hypothesis Testing

➢ The Test of Significance Approach to Hypothesis Testing

- Steps:

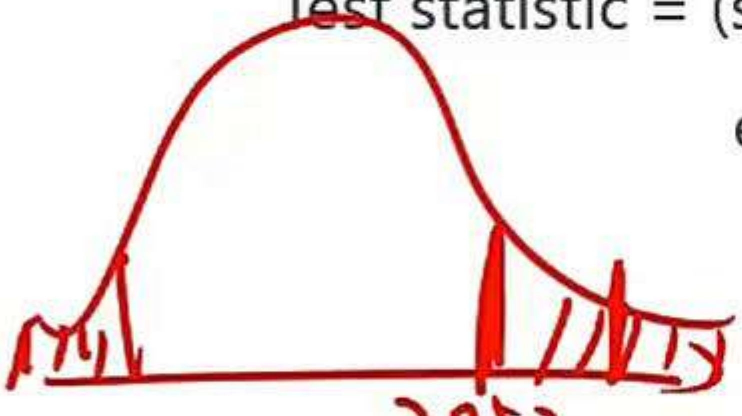
$\mu = 18.5$

① State the null hypothesis H_0 and alternative hypothesis H_1

$$H_0: \mu_X = 18.5, \quad H_1: \mu_X \neq 18.5$$

② Select the test statistic and determine the distribution

Test statistic = (sample statistic – hypothesized value)/(standard error of the sample statistic)



$$t = \frac{(\bar{X} - \mu_X)}{S_X/\sqrt{n}} \sim t_{(n-1)}$$

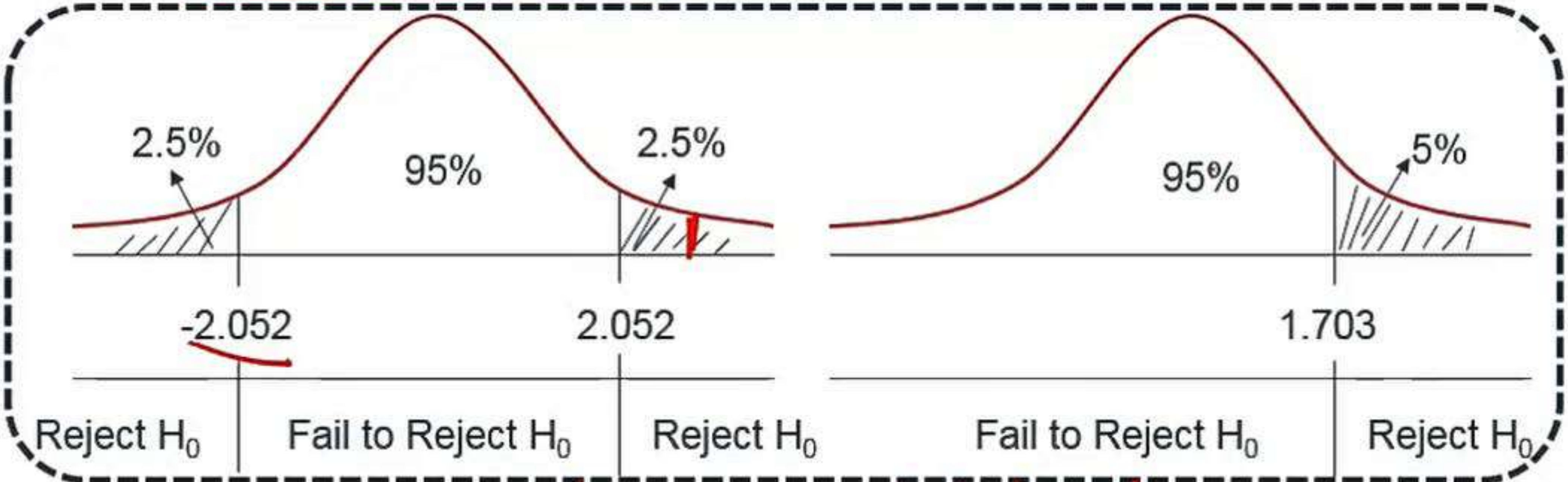
$$\frac{23.25 - 18.5}{9.49/\sqrt{28}} = 2.65$$

③ Choose the level of significance α(5%)

④ Obtain critical t value $t_{\frac{\alpha}{2}(n-1)} = 2.052$

Statistical Inference: Estimation and Hypothesis Testing

⑤ The region of rejection



计算出 查表查出来的

✓ Reject H_0 if |test statistic| > critical value.

✓ Fail to reject H_0 if |test statistic| < critical value accept H_0

✓ We can never say "accept" H_0.

✓ State the conclusion: μ_X is (not) significantly different from μ_0.

μ_x 显著区别于 H_0

Statistical Inference: Estimation and Hypothesis Testing

➢ The Rule of P Value

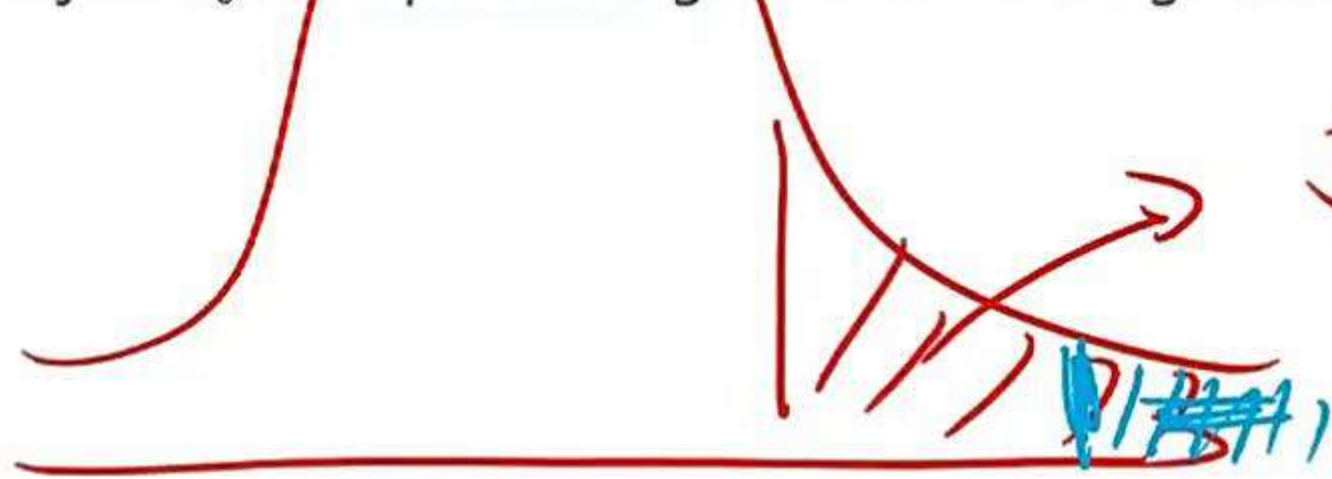
：p-value 越小越拒绝

- P value is the smallest level of significance for which the null hypothesis can be rejected.
- Return to our P/E example,

$$t = \frac{(\bar{X} - \mu_X)}{S_X/\sqrt{n}} = \frac{(23.25 - 18.5)}{9.49/\sqrt{28}} = 2.65 \sim t_{\frac{p}{2}}(27) \to P \approx 0.015$$

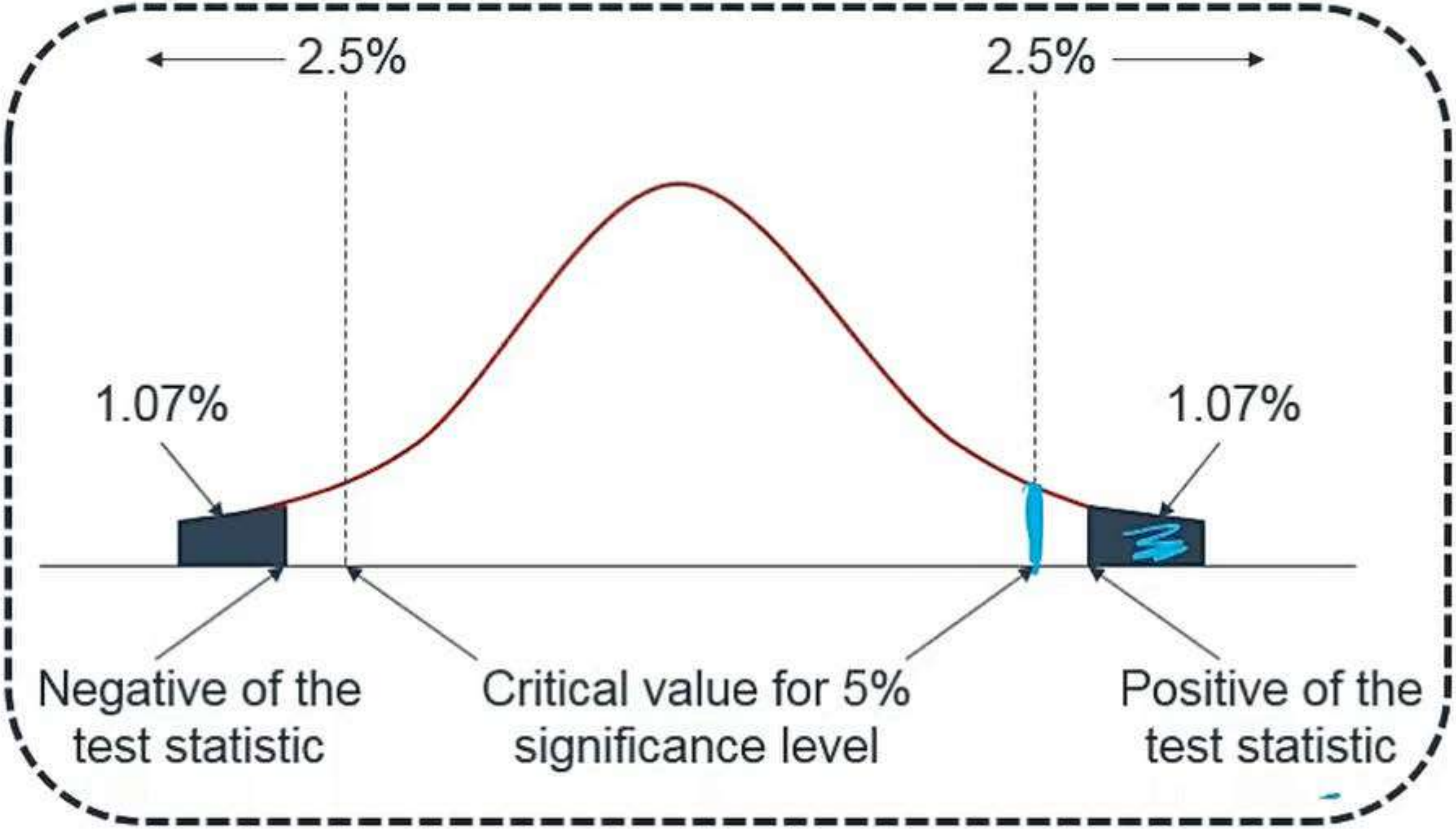
拒

- Reject H_0 if the p-value is less than the significance level of the hypothesis test.
- Do not reject H_0 if the p-value is greater than the significance level.



拒

P-value testing



- If P-value < alpha, we reject null hypothesis.

Test of Single Population Mean

分类

- H_0: $\mu = \mu_0$
  - z-test vs. t-test

	Normal population, n < 30	$n \geq 30$
Known population variance (σ^2)	z-test	z-test
Unknown population variance	t-test	t-test or z-test

  - z-statistic

$$\boxed{z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}}$$

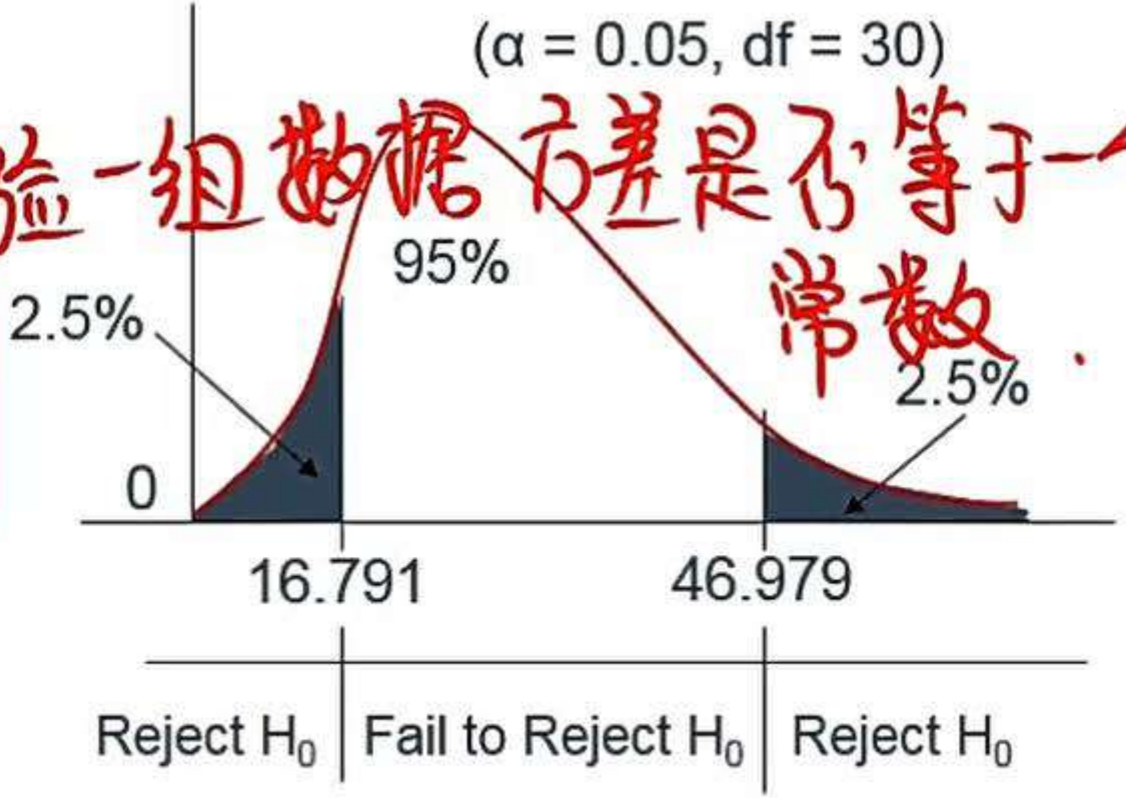
  - t-statistic

$$\boxed{t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}}$$

Test of Single Population Variance

- H_0: $\sigma^2 = \sigma_0^2$
  - The Chi-Square test ：检验一组数据方差是否等于一个常数.

2. $$\boxed{\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad df = n - 1}$$

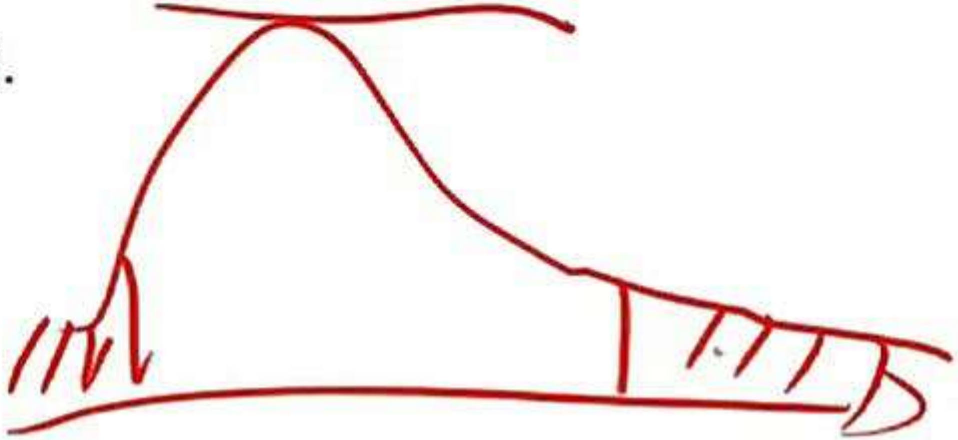


  - Where:
    - ✓ n = sample size
    - ✓ s^2 = sample variance
    - ✓ σ_0^2 = hypothesized value for the population variance

Test of Variances Difference

- **H_0: $\sigma_1^2 = \sigma_2^2$**
  - The F-test：检验两组数据方差是否相等

- $$\boxed{F = \frac{s_1^2}{s_2^2} \quad \mathrm{df}_1 = n_1 - 1 \quad \mathrm{df}_2 = n_2 - 1}$$
  - Always put the larger variance in the numerator ($s_1^2 > s_2^2$).
  - The rejection region is always the right-side tail, no matter the test is one-tailed or two-tailed.

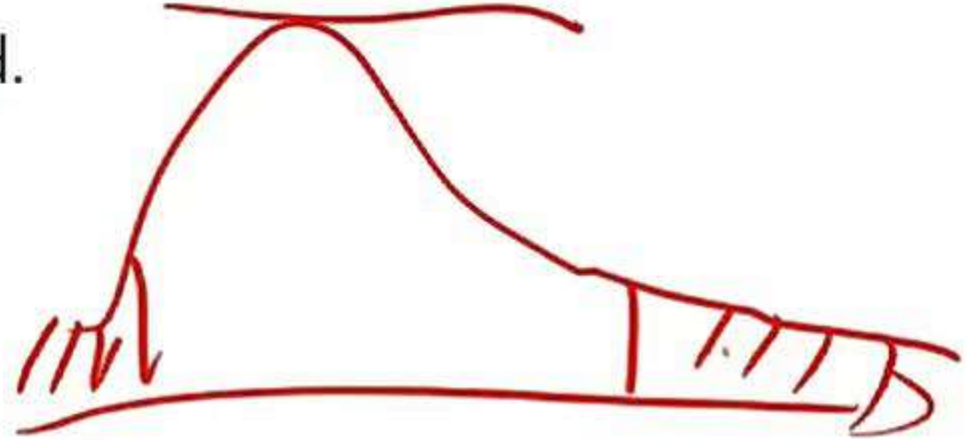


Test of Variances Difference

- $H_0: \sigma_1^{\,2} = \sigma_2^{\,2}$
  - The F-test：检验两组数据方差是否相等

- $$F = \frac{s_1^2}{s_2^2} \quad \mathrm{df}_1 = n_1 - 1 \quad \mathrm{df}_2 = n_2 - 1$$
  - Always put the larger variance in the numerator ($s_1^2 > s_2^2$).
  - The rejection region is always the right-side tail, no matter the test is one-tailed or two-tailed.

"one tailed"

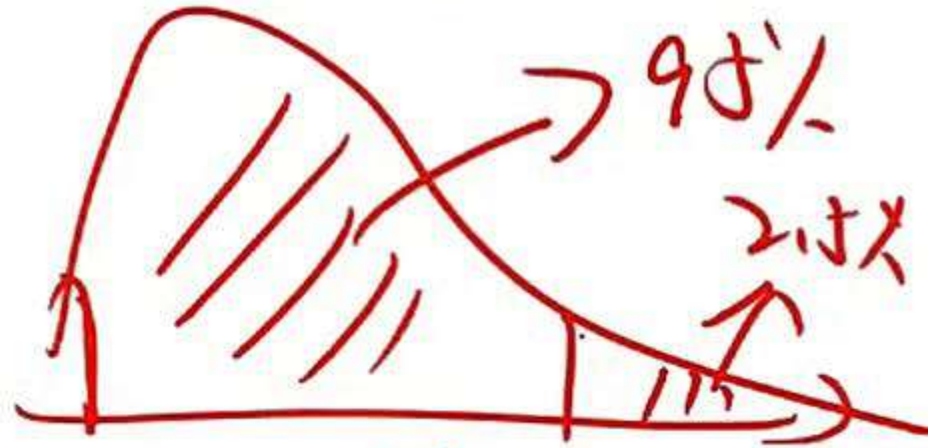


Test of Variances Difference

95%

双尾

- H_0: $\sigma_1^2 = \sigma_2^2$
  - The F-test：检验两组数据方差是否相等

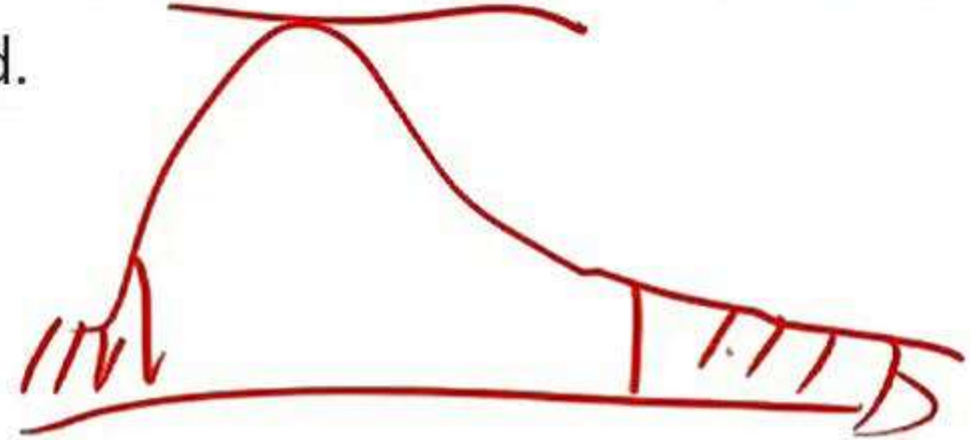


- $$F = \frac{s_1^2}{s_2^2} \quad df_1 = n_1 - 1 \quad df_2 = n_2 - 1$$

α

  - Always put the larger variance in the numerator ($s_1^2 > s_2^2$).
  - The rejection region is always the right-side tail, no matter the test is one-tailed or two-tailed.

"one tailed"



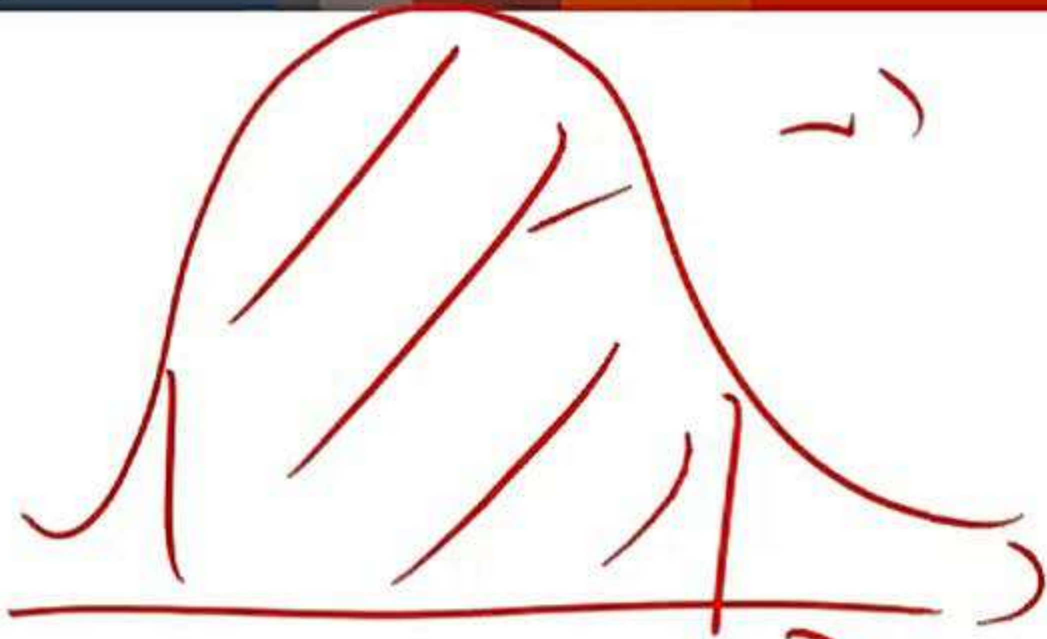
Summary of Hypothesis Testing

Test type	Assumptions	H_0	Test-statistic	distribution
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)
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	$\sigma^2 = \sigma_0^2$	$F = s_1^2/s_2^2$	$F(n_1-1, n_2-1)$

Type I and Type II Errors

- **Type I error**
  - Reject the null hypothesis when it's actually true.
- **Type II error**
  - Fail to reject the null hypothesis when it's actually false.
- **Significance level (α)**
  - The probability of making a Type I error:
    Significance level = P(Type I error)
- **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)

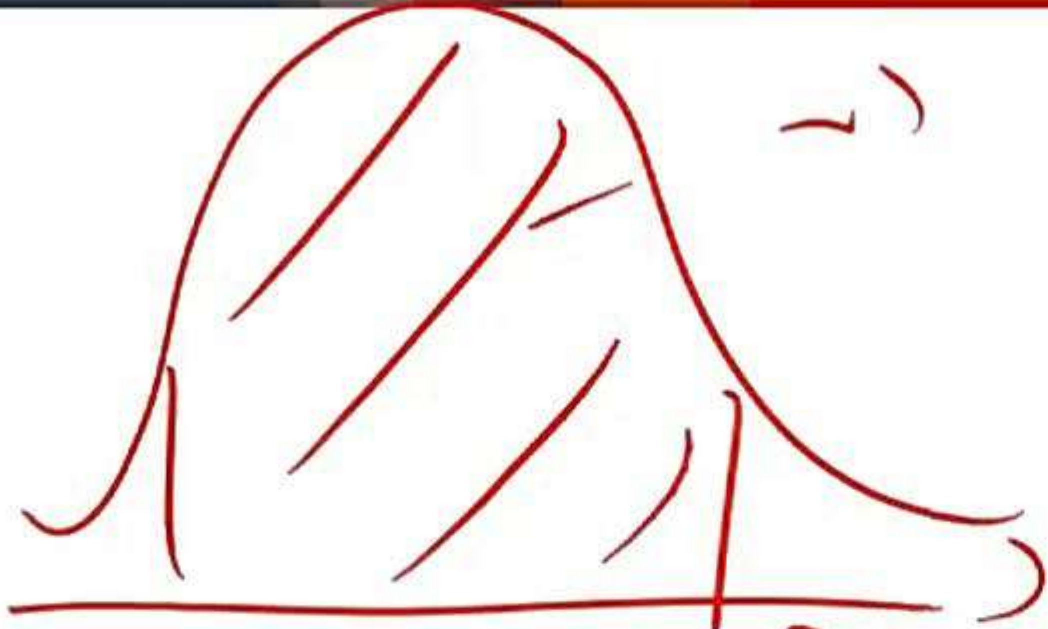


confidence level = P(not reject | H_0: true)

significance level = P(reject | H_0: true).

P(Type II) = P(not reject | H_0: false).

power of test = P (reject | H_0: false)



confidence level $= P(\text{not reject} \mid H_0: \text{true})$

$+$

significance level $= P(\text{reject} \mid H_0: \text{true})$

$P(\text{Type II}) = P(\text{not reject} \mid H_0: \text{false})$

$+$

power of test $= P(\text{reject} \mid H_0: \text{false})$

Type I & Type II 性质：

Type I & Type II 此消彼长.

有伪

$n \uparrow$ P(Type I) & P(Type II) $\downarrow$

Example 2

D

- ~~According to the Basel back testing framework guidelines, penalties start to apply is there are five or more exceptions during the previous year.~~ The Type I error rate of this test is 11 percent. The power of the test is 87 percent. This implies that there is a(an):

  A. 89% probability regulators will reject the correct model. 弃真

  B. 11% probability regulators will reject the incorrect model. ✓

  89% C. 87% probability regulators will not reject the correct model.

  D. 13% probability regulators will not reject the incorrect model.

- Correct Answer：D

$$P(\text{Type I}) = P(\text{reject} \mid H_0:\text{true}) = 11\% \quad \alpha$$

$$P(\text{Type II}) = P(\text{not reject} \mid H_0:\text{false}) = 13\%$$

总结：

一 Term { sample statistic
population parameter
sampling error

二. CLT { 用途：将未知分布转化成正态分布
定义/内容：
性质：mean variance、使用范围

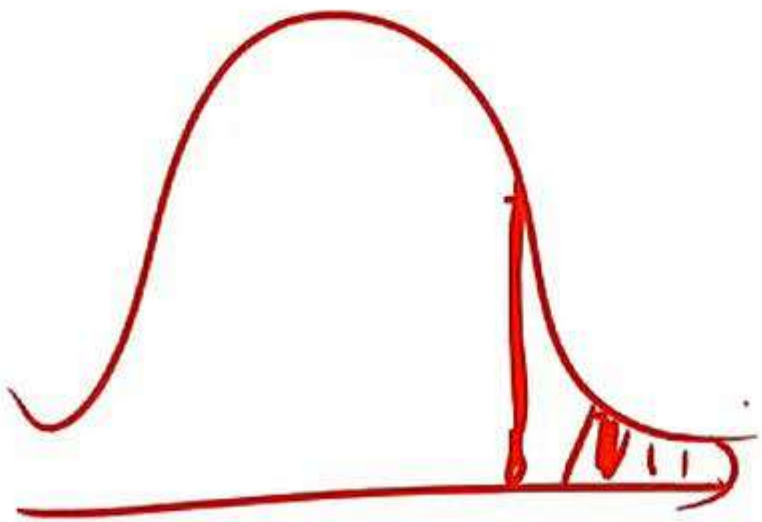
三 点估计 & 区间估计 { 定义
区间估计公式
$\bar{X} \pm k \cdot SE$
z，t 使用情况 {

四 Hypothesis test steps: ① H_0 H_1

② t Z χ^2 F

H_0: $\mu=170$

$\bar{x}=180$ $\quad \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \square$



③ significance level

④ 查 >=< 计

⑤ 得出结论

五、P-value：越小越拒绝

六、Type I / Type II

- 定义.
- confidence level
- power of test $= P(\text{reject} \mid H_0: \text{false})$
- 性质：① 此消彼长
  ② $n \uparrow$ 二者同时 $\downarrow$

Example 1

- Sunstar is a mutual fund with a stated objective of controlling volatility, as measured by the standard deviation of monthly returns. Given the information below, you are asked to test the hypothesis that the volatility of Sunstar's return is equal to 5%. Mean monthly return: 2.5%; Monthly standard deviation: 4.9%; Number of observations: 30.

$n=30$ $H_0: \sigma = 5\%$ $\bar{x}$ $S = 4.9\%$

Chi-square table

df\p	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010
26	12.19815	13.84390	15.37916	17.29188	35.56317	38.88514	41.92317	45.64168
27	12.87850	14.57338	16.15140	18.11390	36.74122	40.11327	43.19451	46.96294
28	13.56471	15.30786	16.92788	18.93924	37.91592	41.33714	44.46079	48.27824
29	14.25645	16.04707	17.70837	19.76774	39.08747	42.55697	45.72229	49.58788
30	14.95346	16.79077	18.49266	20.59923	40.25602	43.77297	46.97924	50.89218

t-table

df\p	0.40	0.25	0.10	0.05	0.025	0.01
26	0.25596	0.68404	1.31497	1.70562	2.05553	2.47863
27	0.25586	0.68369	1.31370	1.70329	2.05183	2.47266
28	0.25577	0.68335	1.31253	1.70113	2.04841	2.46714
29	0.25568	0.68304	1.31143	1.69913	2.04523	2.46202
30	0.25561	0.68276	1.31042	1.69726	2.04227	2.45726

- What is the correct test should be used and what is the correct conclusion at 5% level of significance?
  - A. Chi-square test, reject the hypothesis that volatility is 5%.
  - B. Chi-square test, do not reject the hypothesis that volatility is 5%.
  - C. t test, reject the hypothesis that volatility is 5%.
  - D. t test, do not reject the hypothesis that volatility is 5%.
- Correct Answer : B

Example 1

- Sunstar is a mutual fund with a stated objective of controlling volatility as measured by the standard deviation of monthly returns. Given the information below, you are asked to test the hypothesis that the volatility of Sunstar's return is equal to 5%. Mean monthly return: 2.5%; Monthly standard deviation: 4.9%; Number of observations: 30.

$n=30$ ① $H_0: \sigma = 5\%$ $\bar{x}$ $S = 4.9\%$

Chi-square table

df\p	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010
26	12.19815	13.84390	15.37916	17.29188	35.56317	38.88514	41.92317	45.64168
27	12.87850	14.57338	16.15140	18.11390	36.74122	40.11327	43.19451	46.96294
28	13.56471	15.30786	16.92788	18.93924	37.91592	41.33714	44.46079	48.27824
29	14.25645	16.04707	17.70837	19.76774	39.08747	42.55697	45.72229	49.58788
30	14.95346	16.79077	18.49266	20.59923	40.25602	43.77297	46.97924	50.89218

② $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{29 \times 4.9\%^2}{5\%^2} = 27.8516$

t-table

df\p	0.40	0.25	0.10	0.05	0.025	0.01
26	0.25596	0.68404	1.31497	1.70562	2.05553	2.47863
27	0.25586	0.68369	1.31370	1.70329	2.05183	2.47266
28	0.25577	0.68335	1.31253	1.70113	2.04841	2.46714
29	0.25568	0.68304	1.31143	1.69913	2.04523	2.46202
30	0.25561	0.68276	1.31042	1.69726	2.04227	2.45726

③ 5% ④ 45.72229

Example 1

- What is the correct test should be used and what is the correct conclusion at 5% level of significance?

  A. Chi-square test, reject the hypothesis that volatility is 5%.

  B. Chi-square test, do not reject the hypothesis that volatility is 5%.

  C. t test, reject the hypothesis that volatility is 5%.

  D. t test, do not reject the hypothesis that volatility is 5%.

- Correct Answer : B

Example 3

- Which of the following statements regarding hypothesis testing is incorrect?

   A. Type II error refers to the failure to reject the null hypothesis when it is actually false.

   B. Hypothesis testing is used to make inferences about the parameters of a given population on the basis of statistics computed for a sample that is drawn from that population.

   C. All else being equal, the decrease in the chance of making a Type I error comes at the cost of increasing the probability of making a Type II error.

   D. The p-value decision rule is to reject the null hypothesis if the p-value is greater than the significance level.

- Correct Answer : D

Example 3

- Which of the following statements regarding hypothesis testing is incorrect?

D

A. Type II error refers to the failure to reject the null hypothesis when it is actually false. 存伪

B. Hypothesis testing is used to make inferences about the parameters of a given population on the basis of statistics computed for a sample that is drawn from that population.

C. All else being equal, the decrease in the chance of making a Type I error comes at the cost of increasing the probability of making a Type II error.

D. The p-value decision rule is to reject the null hypothesis if the p-value is greater than the significance level.

Smaller

- Correct Answer : D

Example 4

- What does a hypothesis test at the 5% significance level mean?
  - A. P(not reject H_0 | H_0 is true) = 0.05
  - B. P(not reject H_0 | H_0 is false) = 0.05
  - C. P(reject H_0 | H_0 is true) = 0.05
  - D. P(reject H_0 | H_0 is false) = 0.05

C

P(弃真) = 5%

P(reject | H_0: true) = 5%

- Correct Answer : C

~~确定函数关系~~ 工资 100 × 8

统计相关关系：

Linear Regression-Linear Regression with One Regressor

一、Term

二、OLS：最小二乘法

三、Assumptions

四 ANOVA Table

五、对于回归系数的检验.

Dependent and Independent Variable

➢ Regression analysis

- Regression analysis is concerned with the study of the relationship between one variable called the dependent or explained variable and one or more other variables called independent or explanatory variables.

Dependent variable 依赖变量/因变量/被解释变量

$$\boxed{Y = b_0 + b_1 X + \varepsilon}$$

Independent variable 独立/自/解释

- The objectives of regression analysis: to predict or forecast dependent variable.

Interpretation of Regression Coefficients

- **Interpretation of regression function**

  斜率项

  - An estimated slope coefficient of 2 would indicate that the dependent variable will change two units for every 1 unit change in the independent variable.
  - The intercept term of 2% can be interpreted to mean that the independent variable is zero, the dependent variable is 2%.

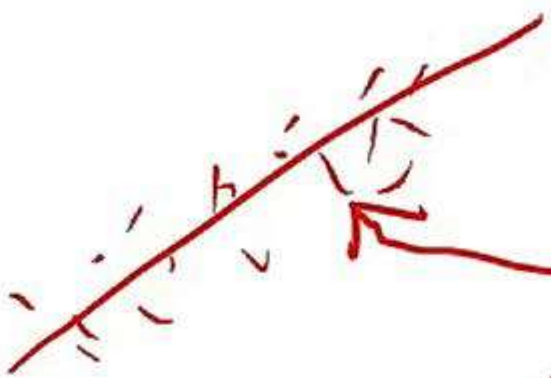
  截距项

$$Y = 2 + 3X$$

Interpretation of Regression Coefficients

- **Interpretation of regression coefficients**
  - The estimated intercept coefficient ($\hat{b}_0$) is interpreted as the value of Y when X is equal to zero.
  - The estimated slope coefficient ($\hat{b}_1$) defines the sensitivity of Y to a change in X. The estimated slope coefficient ($\hat{b}_1$) equals covariance divided by variance of X.

$\hat{Y} = \hat{b}_0 + \hat{b}_1 X$



Slope coefficient

$$Y = b_0 + b_1 X + \varepsilon$$

Intercept coefficient

Error term

Ordinary Least Squares (OLS) 最小二乘法

- **Ordinary least squares (OLS)**：计算 $\hat{b_0}$ $\hat{b_1}$
  - The OLS estimation is the process of estimating the population parameter b_i using the corresponding b_i value, which minimizes the square residual (i.e., the error terms).
  - The OLS sample coefficients are those that:

$$\text{minimize} \sum \varepsilon_i^2 = \sum [Y_i - (\hat{b}_0 + \hat{b}_1 \times X_1)]^2$$

$$\hat{b}_1 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})2} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1\bar{X}$$

  - The estimated intercept coefficient ($\hat{b}_0$): the point ($\bar{X}, \bar{Y}$) is on the regression line.

Ordinary Least Squares (OLS) 最小二乘法

- **Ordinary least squares (OLS)**：计算 $\hat{b}_0$ $\hat{b}_1$
  - The OLS estimation is the process of estimating the population parameter b_i using the corresponding b_i value, which minimizes the square residual (i.e., the error terms).
  - The OLS sample coefficients are those that:

$$Y = \hat{b}_0 + \hat{b}_1 X$$

$$\bar{Y} = \hat{b}_0 + \hat{b}_1 \bar{X}$$

$$\text{minimize} \sum \varepsilon_i^2 = \sum [Y_i - (\hat{b}_0 + \hat{b}_1 \times X_1)]^2$$

$$\hat{b}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})2} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{\rho \sigma_X \sigma_Y}{\sigma_X^2} = \rho \cdot \frac{\sigma_Y}{\sigma_X}$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$$

  - The estimated intercept coefficient ($\hat{b}_0$): the point ($\bar{X}, \bar{Y}$) is on the regression line.

Ordinary Least Squares Regression

➢ **Example: Sample of Returns and Corresponding Lockup periods**

	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	Cov(X,Y)	Var(X)	e_i	$\sum(Y_i - \hat{Y})^2$	$\sum(Y_i - \bar{Y})^2$	$\hat{Y}_i$	$\sum(\hat{Y}_i - \bar{Y})^2$
	5	10	-2.5	-6	15	6.25	-1	1	36	11	25
	6	12	-1.5	-4	6	2.25	-1	1	16	13	9
	7	19	-0.5	3	-1.5	0.25	4	16	9	15	1
	8	16	0.5	0	0	0.25	-1	1	0	17	1
	9	18	1.5	2	3	2.25	-1	1	4	19	9
	10	21	2.5	5	12.5	6.25	0	0	25	21	25
sum	**45**	**96**	**0**	**0**	**35**	**17.5**	**0**	**20**	**90**	**96**	**70**
$\bar{X}$ $\bar{Y}$ average	**7.5**	**16**									

- $b_1 = 35/17.5 = 2$; $b_0 = 16 - 2\times7.5 = 1$;
- The sample regression function is: $Y_i = 1 + 2\times X_i + e_i$.
- According to the data, on average a hedge fund with a lockup period of 6 years will have a 2% higher return than a hedge fund with a 5-year lockup period.

The Basics of Simple Linear Regression

假设

- **The assumptions of simple linear regression**
  - X and Y have a linear relationship.
  - X is not random, and the condition that X is uncorrelated with the error term can substitute the condition that X is not random.
  - The expected value of the error term is zero (i.e., $E(\varepsilon_i) = 0$).
  - 次 The variance of the error term is constant (i.e., the error terms are homoskedastic).
  - The error term is uncorrelated across observations (i.e., $E(\varepsilon_i \varepsilon_j) = 0$ for all $i \neq j$).
  - The error term follows normal distribution.

$$Y = b_0 + b_1 X + \varepsilon$$

The Basics of Simple Linear Regression

假设

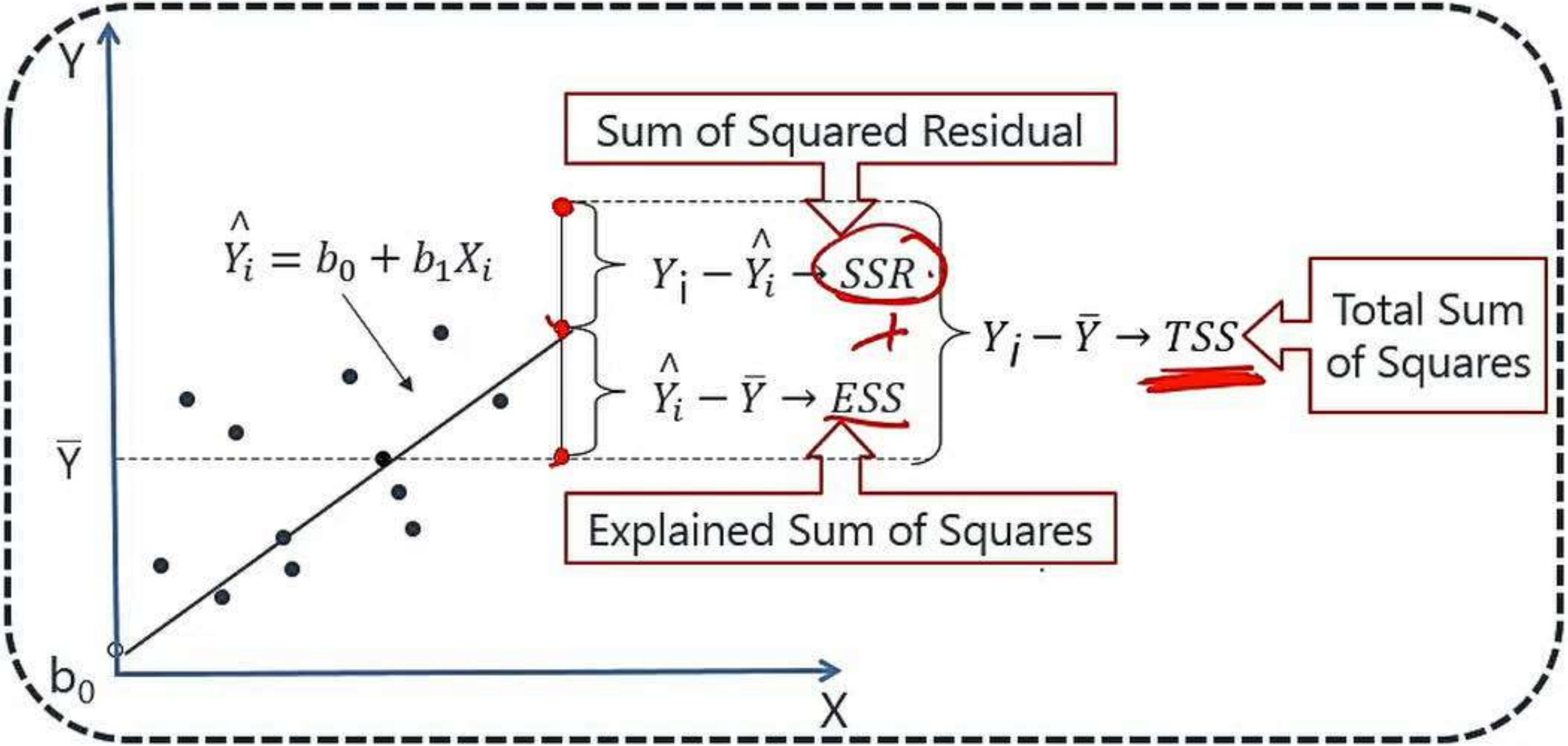
The assumptions of simple linear regression

- X and Y have a linear relationship.
- X is not random, and the condition that X is uncorrelated with the error term can substitute the condition that X is not random.
- The expected value of the error term is zero (i.e., $E(\varepsilon_i) = 0$).
- The variance of the error term is constant (i.e., the error terms are homoskedastic). 同质性
- The error term is uncorrelated across observations (i.e., $E(\varepsilon_i\varepsilon_j) = 0$ for all $i \neq j$). ε_i ε_j
- The error term follows normal distribution. 高斯

$$Y = b_0 + b_1 X + \varepsilon$$

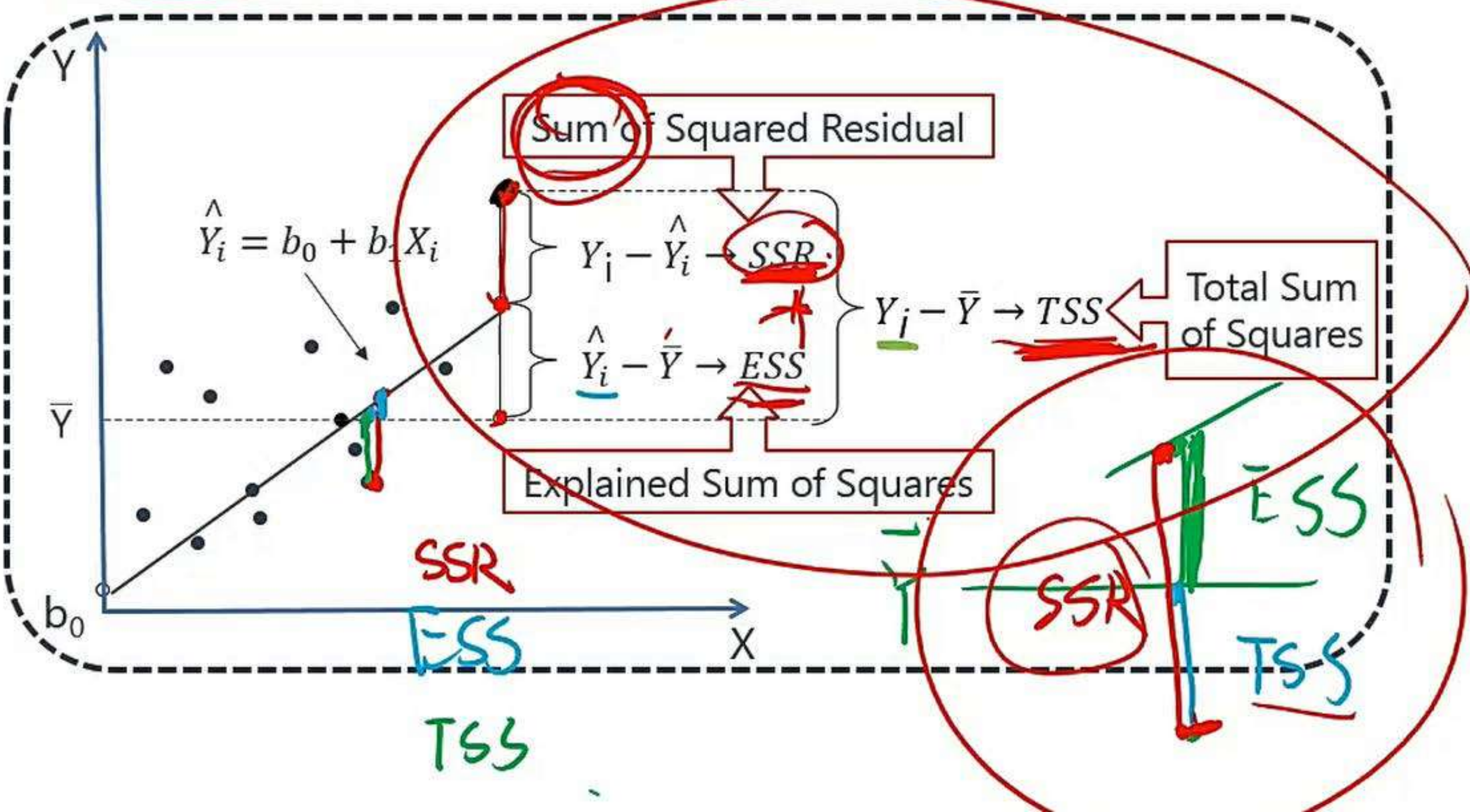
Analysis of Variance (ANOVA) Table

- **Components**



Analysis of Variance (ANOVA) Table

➢ **Components**



The Standard Error of the Regression

- Total sum of squares=explained sum of squares+sum of squared residuals

$$\sum(Yi - \bar{Y})^2 \quad = \quad \sum(\hat{Y}_i - \bar{Y})^2 \quad + \quad \sum(Yi - \hat{Y}_i)^2$$

$$TSS \quad = \quad ESS \quad + \quad SSR$$

- SER is a measure of the spread of the observations around the regression line, measured in the units of the dependent variable.

$$\boxed{SER = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}}}$$

- The **standard error of the regression (SER)** is an estimator of the standard deviation of the regression error u_i
- The SER measures the "fit" of the regression line. The smaller the standard error, the better it fits. if the relationship is very strong, SER will be low. (relative to total variability)

Analysis of Variance (ANOVA) Table

- **Analysis of Variance (ANOVA) Table**

	df	SS	MSS
Regression	k = 1	ESS	ESS/k
Residual	n – 2	SSR	SSR/(n – 2)
Total	n – 1	TSS	-

- **Notes**
  - Total sum of squares (TSS) is also known as sum of squares total (SST).
  - Explained sum of squares (ESS) is also known as regression sum of squares (RSS).
  - Sum of squares residual (SSR) is also known as sum of squares errors (SSE).
  - Standard error of regression (SER) is also known as standard error of estimate (SEE).

The Standard Error of the Regression

- Total sum of squares=explained sum of squares+sum of squared residuals

$$\sum(Yi-\bar{Y})^2 = \sum(\hat{Y}_i-\bar{Y})^2 + \sum(Yi-\hat{Y}_i)^2$$

$$TSS = ESS + SSR$$

- SER is a measure of the spread of the observations around the regression line, measured in the units of the dependent variable.

$$\boxed{SER=\sqrt{\frac{\sum(Y_i-\hat{Y}_i)^2}{n-2}}} \quad \sqrt{\frac{\Sigma(\hat{Y}_i-\hat{Y}_i)^2}{n-k-1}}$$

- The **standard error of the regression (SER)** is an estimator of the standard deviation of the regression error u_i
- The SER measures the "fit" of the regression line. The smaller the standard error, the better it fits. if the relationship is very strong, SER will be low. (relative to total variability)

The Standard Error of the Regression

- Total sum of squares=explained sum of squares+sum of squared residuals

$$\sum(Yi-\bar{Y})^2 = \sum(\hat{Y}_i-\bar{Y})^2 + \sum(Yi-\hat{Y}_i)^2$$

$$TSS = ESS + SSR$$

- SER is a measure of the spread of the observations around the regression line, measured in the units of the dependent variable.

$$SER=\sqrt{\frac{SSR}{n-k-1}}$$

$$SER=\sqrt{\frac{\sum(Y_i-\hat{Y}_i)^2}{n-2}}$$

$$\sqrt{\frac{\Sigma(Y_i-\hat{Y}_i)^2}{n-k-1}}$$

- The **standard error of the regression (SER)** is an estimator of the standard deviation of the regression error u_i
- The SER measures the "fit" of the regression line. The smaller the standard error, the better it fits. if the relationship is very strong, SER will be low. (relative to total variability)

Analysis of Variance (ANOVA) Table

marginal

➢ Analysis of Variance (ANOVA) Table

	df	SS	MSS
Regression	k = 1	ESS	ESS/k
Residual	n – 2 n−k−1	SSR	SSR/(n – 2)
Total	n – 1	TSS	-

➢ **Notes**

Total

- Total sum of squares (TSS) is also known as sum of squares total (SST).
- Explained sum of squares (ESS) is also known as regression sum of squares (RSS). ESSR E R
R E
- Sum of squares residual (SSR) is also known as sum of squares errors (SSE).
- Standard error of regression (SER) is also known as standard error of estimate (SEE).

Measures of Fit

$Y = b_0 + b_1 X_1 + b_2 X_2$

ρ

- **R^2 (the Coefficient of Determination)** 决定系数
  - A measure of the "goodness of fit" of the regression. It is interpreted as a percentage of variation in the dependent variable explained by the independent variable. Its limits are $0 \leq R^2 \leq 1$.

$$\boxed{R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}}$$

$SSR \rightarrow$ residual

  - In a simple two-variable regression, the square root of R^2 is the correlation coefficient (r) between X_i and Y_i.

※考点 $\boxed{r^2 = R^2 \rightarrow r = \pm\sqrt{R^2}}$ 一元线性回归中

ρ ※

$Y = b_0 + b_1 X_1$ ρ

Measures of Fit

> **The Different between the R^2 and Correlation Coefficient** ρ

- The correlation coefficient indicates the sign of the relationship between two variables, whereas the coefficient of determination does not.
- The coefficient of determination can apply to an equation with several independent variables, and it implies a explanatory power, while the correlation coefficient only applies to two variables and does not imply explanatory between the variables.

ρ：衡量两个变量之间的相关关系

R^2：衡量模型的解释力度.

Confidence Interval

➢ The confidence interval for the regression coefficient, b_1, is calculated as:

$\bar{X} \pm K \cdot SE$

$$\hat{b}_1 \pm (t_c \times s_{\hat{b}_1}), or \ [\hat{b}_1 - (t_c \times s_{\hat{b}_1}) < b_1 < \hat{b}_1 + (t_c \times s_{\hat{b}_1})]$$

- t_c: The critical two-tailed t value of the selected confidence level with an appropriate number of degrees of freedom is equal to the number of sample observations minus 2 (i.e., n-2).
- S_{b1}:The standard error of the regression coefficient.

➢ $S_{\hat{b}_1}$ is the function of SER: As the SER rises, $S_{\hat{b}_1}$ also increases, and the confidence interval is widened.

Hypothesis Testing

➢ **Regression Coefficient Hypothesis Testing**

- The hypothesis that if the true slope is zero ($b_1 = 0$). The appropriate test structure for zero and alternative assumptions is:

$$H_0: b_1 = 0 \quad H_a: b_1 \neq 0$$

- A t-test may also be used to test the hypothesis that the true slope coefficient b_1, is equal to some hypothesized value. Letting $\hat{b}_1$ be the point estimate for b_1, the appropriate test statistic with $n - 2$ degrees of freedom is:

$$\boxed{t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} \sim t(n-2)}$$

- The decision rule for tests of significance for regression coefficients is:

$$\text{Reject } H_0 \text{ if } t > +t_{critical} \text{ or } t < -t_{critical}$$

Hypothesis Testing

➢ **Regression Coefficient Hypothesis Testing**

- The hypothesis that if the true slope is zero ($b_1 = 0$). The appropriate test structure for zero and alternative assumptions is:

① H_0: $b_1 = 0$ H_a: $b_1 \neq 0$

- A t-test may also be used to test the hypothesis that the true slope coefficient b_1, is equal to some hypothesized value. Letting $\hat{b}_1$ be the point estimate for b_1, the appropriate test statistic with n – 2 degrees of freedom is:

② $$\boxed{t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} \sim t(n-2)}$$ n-k-1

- The decision rule for tests of significance for regression coefficients is:

④ Reject H_0 if $t > +t_{critical}$ or $t < -t_{critical}$ ③

Example

- Calculating the confidence interval for a regression coefficient
  - The estimated slope coefficient, B_1, from WPO regression is 0.64 with a standard error equal to 0.26. Assuming that the sample had 36 observations, calculate the 95% confidence interval for B_1.

$\bar{X} \pm k \cdot SE$

$0.64 \pm 2.03 \times 0.26$

- Answer :
  - The confidence interval at 95% for b_1 is:

$$0.64 \pm (2.03)(0.26) = 0.64 \pm 0.53 = 0.11 \text{ to } 1.17$$

  - Because this confidence interval does not include 0, we can conclude that the slope coefficient is significantly different from 0.

$$t = \frac{b_1 - B_1}{s_{b_1}} = \frac{0.64 - 0}{0.26} = 2.46$$

  - Because $t > t_{critical}$ (i.e., 2.46 > 2.03), we reject the null hypothesis and conclude that the slope is different from zero.

- Calculating the confidence interval for a regression coefficient
  - The estimated slope coefficient, B_1, from WPO regression is 0.64 with a standard error equal to 0.26. Assuming that the sample had 36 observations, calculate the 95% confidence interval for B_1.

$\bar{X} \pm k \cdot SE$

$0.64 \pm 2.03 \times 0.26$

- Answer :
  - The confidence interval at 95% for b_1 is:

$$0.64 \pm (2.03)(0.26) = 0.64 \pm 0.53 = 0.11 \text{ to } 1.17$$

  - Because this confidence interval does not include 0, we can conclude that the slope coefficient is significantly different from 0.

$$t = \frac{\hat{b}_1 - B_1}{s_{b_1}} = \frac{0.64 - 0}{0.26} = 2.46$$

  - Because $t > t_{critical}$ (i.e., 2.46 > 2.03), we reject the null hypothesis and conclude that the slope is different from zero.

总结：

一、Term：$\underline{Y} = \underline{b_0} + \underline{b_1} \underline{X} + \underline{\varepsilon}$. 一元线性

$\hat{Y} = \hat{b_0} + \hat{b_1} X$

二、OLS：用途：$\hat{b_0}$ $\hat{b_1}$

计算公式：

三、假设：6个.

四、ANOVA Table

衡量线性回归拟合程度

$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$

在一元中 $R^2 = \rho^2$.

五、线性回归的假设检验：五个步骤

The Basics of Multiple Regression

- Multiple regression is regression analysis with more than one independent variable.
  - The multiple linear regression model

$$\boxed{Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki} + \varepsilon_i}$$

  - Predicted value of the dependent variable

$$\boxed{\hat{Y} = \hat{b}_0 + \hat{b}_1 \hat{X}_1 + \hat{b}_2 \hat{X}_2 + \cdots + \hat{b}_k \hat{X}_k}$$

  - OLS Estimator
  - minimize $\sum \varepsilon_i^2 = \sum [Y_i - (b_0 + \sum_{i=1}^{n} b_i \times X_i)]^2$

Interpreting the Multiple Regression Results

- When independent variables are all equal to zero, dependent variable equals to the intercept term b_0
- The slope b_1 is the change in dependent variable associated with a unit change in independent variable, keeping other independent variables unchanged. This is why the slope coefficients in the multiple regression are sometimes referred to as local slope coefficients.

Multiple Regression Assumptions

- **The assumptions of the multiple linear regression**
  - There is a linear relationship between dependent and independent variables.
  - The independent variable is not random (Or X is not related to the error item). There is no precise linear relationship between any two or more independent variables.
  - The expected value of the error term is zero (i.e., $E(\varepsilon_i)=0$) .
  - The variance of the error term is constant (i.e., the error terms are homoskedastic).
  - The error term is uncorrelated across observations (i.e., $E(\varepsilon_i\varepsilon_j)=0$ for all $i \neq j$). $\rho(\varepsilon_i, \varepsilon_j) = 0$
  - The error term is normally distributed.

Analysis of Variance (ANOVA) Table

➢ **Analysis of Variance (ANOVA) Table**

	df	SS	MSS
Regression	k	ESS	ESS/k
Residual	n – k – 1	SSR	SSR/(n – k – 1)
Total	n – 1	TSS	-

➢ **Specially, in the single regression, the ANOVA table is:**

	df	SS	MSS
Regression	k = 1	ESS	ESS/k
Residual	n – 2	SSR	SSR/(n – 2)
Total	n – 1	TSS	-

R^2 and Adjusted R^2

- **Standard Error of Regression (SER)**

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

Σ

- **R^2 and Adjusted R^2**
  - In multiple regression, the R^2 increases whenever a regressor (independent variable) is added, unless the estimated efficient on the added regressor is exactly zero.
  - The adjusted R^2 is a modified version of the R^2 that does not necessarily increase with a new independent variable is added. Adjusted R^2 is given by:

$$\text{Adjusted } R^2 = 1 - \frac{SSR/n-k-1}{TSS/n-1} = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$$

✓ Adjusted $R^2 \leq R^2$; adjusted R^2 may be less than zero.

R^2 and Adjusted R^2

➢ **Standard Error of Regression (SER).**

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

Σ

➢ **R^2 and Adjusted R^2**

- In multiple regression, the R^2 increases whenever a regressor (independent variable) is added, unless the estimated efficient on the added regressor is exactly zero.
- The adjusted R^2 is a modified version of the R^2 that does not necessarily increase with a new independent variable is added. Adjusted R^2 is given by:

residual

$$\text{Adjusted } R^2 = 1 - \frac{SSR/n-k-1}{TSS/n-1} = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$$

$(1-R^2)$

✓ Adjusted $R^2 \le R^2$; adjusted R^2 may be less than zero.

R^2 and Adjusted R^2

- **The reason of adjusted R^2**
  - To further analyze the importance of an added variable to a regression, we can compute an adjusted R^2.
    - ✓ Mathematically, if the variable with any explanatory power is added to the regression, the determined coefficient (R^2) **will increase**, <u>even if the marginal contribution of the new variable is not statistically significant</u>.
    - ✓ A relatively high R^2 may reflect the influence of a large set of independent variables, rather than the extent to which the set interprets the dependent variable.
    - ✓ This phenomenon is often referred to as overestimate regression.

Example

- Calculating R^2 and adjusted R^2
  - The analyst runs a regression of monthly return on the five independent variables within 60 months. The sum of squares is 570, the sum of the squared errors is 180. Calculate R^2 and adjust the R^2.
  - Assuming that the analyst now adds four independent variables for the regression, R^2 increases to 70.0%. Identify the models that analysts are most likely to like.
- Answer:

$$R^2 = \frac{570 - 180}{570} = 68.42\%$$

$$R_a^2 = 1 - \left(\frac{60 - 1}{60 - 5 - 1}\right)(1 - R^2) = 65.5\%$$

$$R'^2_a = 1 - \left(\frac{60 - 1}{60 - 9 - 1}\right)(1 - R'^2) = 64.6\%$$

Example

➢ Calculating R^2 and adjusted R^2

K = 5

n = 60

- The analyst runs a regression of monthly return on the five independent variables within 60 months. The sum of squares is 570,
12 the sum of the squared errors is 180. Calculate R^2 and adjust the R^2.

K = 9

- Assuming that the analyst now adds four independent variables for the regression, R^2 increases to 70.0%. Identify the models that analysts are most likely to like.

➢ Answer:

$$R^2 = \frac{570 - 180}{570} = 68.42\%$$

$$R_a^2 = 1 - \left(\frac{60 - 1}{60 - 5 - 1}\right)(1 - R^2) = 65.5\%$$

$$R'^2_a = 1 - \left(\frac{60 - 1}{60 - 9 - 1}\right)(1 - R'^2) = 64.6\%$$

Interpreting Regression Results – ANOVA

Figure 2: ANOVA Table

R-squared	0.9341
Adj R-squared	0.8902
Standard error	1.439
Observations	6

	Degrees of Freedom	SS	MSS	F
Explained	2	88.064	44.032	21.261
Residual	3	6.213	2.071	
Total	5	94.277		

Variables	Coeff	Std. Error	t-stat	P-value	Lower 95%	Upper 95%
Intercept	-4.176	3.299	-1.266	0.270	-14.6734	6.3214
Lockup	2.375	0.337	7.047	0.009	1.3003	3.45
Experience	1.986	0.754	2.634	0.076	-0.4132	4.3853

Interpreting Regression Results – ANOVA

Figure 2: ANOVA Table

$R^2 = \frac{ESS}{TSS}$

R-squared	0.9341
Adj R-squared	0.8902
SER Standard error	1.439
Observations n	6

	Degrees of Freedom	SS	MSS	F
Explained	2	88.064	44.032	21.261
Residual	3	6.213	2.071	
Total	5	94.277		

Variables	系数 Coeff	Std. Error	t-stat	P-value	Lower 95%	Upper 95%
Intercept	-4.176	3.299	-1.266	0.270	-14.6734	6.3214
Lockup	2.375	0.337	7.047	0.009	1.3003	3.45
Experience	1.986	0.754	2.634	0.076	-0.4132	4.3853

Interpreting Regression Results – ANOVA

Figure 2: ANOVA Table

R-squared	0.9341
Adj R-squared	0.8902
SER Standard error	1.439
Observations n	6

$$Adj\ R^2 = 1 - \frac{SSR/n-k-1}{TSS/n-1}$$

$$\frac{2.071}{94.277/5}$$

$$\sqrt{\frac{SSR}{n-k-1}}$$

	Degrees of Freedom	SS	MSS	F
Explained	2	88.064	44.032	21.261
Residual	3	6.213	2.071	
Total	n-1 5	94.277		

Variables	Coeff 系数	Std. Error	t-stat	P-value	Lower 95%	Upper 95%
Intercept	-4.176	3.299	-1.266	0.270	-14.6734	6.3214
Lockup	2.375	0.337	7.047	0.009	1.3003	3.45
Experience	1.986	0.754	2.634	0.076	-0.4132	4.3853

Joint Hypothesis Testing

- **Joint Hypothesis Testing**
  - An F-test is used to test whether at least one slope coefficient is significantly different from zero.

H_0: $b_1 = b_2 = b_3 = \ldots = b_k = 0$; H_a: at least one $b_j \neq 0$ ($j = 1$ to k)

- **F-Statistic**

$$\boxed{F = \frac{ESS/k}{SSR/(n-k-1)}}$$

  - The F-test here is always a one-tailed test.
  - The test assesses the effectiveness of the model as a whole in explaining the dependent variable.
  - Decision rule: reject H_0, if $F_{(\text{test-statistic})} > F_{c(\text{critical value})}$.

Interpreting Regression Results – ANOVA

$\bar{X} \pm k \cdot SE$

Figure 2: ANOVA Table

R-squared	0.9341
Adj R-squared	0.8902
SER Standard error	1.439
Observations n	6

	Degrees of Freedom	SS	MSS	F
Explained	k 2	88.064	44.032	21.261
Residual	n-k-1 3	6.213	2.071	
Total	n-1 5	94.277		

F统计量

$\frac{\hat{b}_1 - b_1}{SE}$

Variables	Coeff	Std. Error	t-stat	P-value	Lower 95%	Upper 95%
Intercept	-4.176	3.299	-1.266	0.270	-14.6734	6.3214
Lockup	2.375	0.337	7.047	0.009	1.3003	3.45
Experience	1.986	0.754	2.634	0.076	-0.4132	4.3853

Hypothesis test for a Partial Slope Coefficient

Hypothesis test for a Partial Slope Coefficient

- H_0: $b_j = 0$ ($j = 1, 2, \ldots, k$)

$$\boxed{t = \frac{\hat{b}_j}{s_{\hat{b}_j}} \sim t(n-k-1)}$$

- Regression coefficient confidence interval

$$\boxed{\hat{b}_j \pm (\mathrm{tc} \times s_{\hat{b}_j})}$$

Hypothesis test for a Partial Slope Coeffici

Hypothesis test for a Partial Slope Coefficient

- $H_0: b_j = 0\ (j = 1, 2, \ldots, k)$

① ②
$$\boxed{t = \frac{\hat{b}_j}{s_{\hat{b}_j}} \sim t(n-k-1)}$$

$$\frac{\hat{b}_j - \cancel{b_j}}{S_{\hat{b}_j}}$$

- Regression coefficient confidence interval

$$\boxed{\hat{b}_j \pm (\text{tc} \times s_{\hat{b}_j})}$$

$$\bar{x} \pm k\, SE$$

Hypothesis Testing of Regression Coefficients

➢ **Example:** 10% significance level, 40 observations.

	Coefficient	Standard Error	t-statistic	p-value
Intercept	-10.60%	1.542%	6.87	<0.0001
PR	0.35	0.014	25	<0.0001
YCS	0.21	0.38	0.55	

- (1)

$H_0: PR = 0$
$H_1: PR \neq 0$

$t = \frac{0.35}{0.014}$

(2)

$H_0: PR = 0.2$
$H_1: PR \neq 0.2$

$t = \frac{0.35 - 0.2}{0.014}$

(3)

H_0: intercept $\geq -10.0\%$
H_1: intercept $< -10.0\%$

$t = \frac{-10.6\% - (-10\%)}{1.542\%}$

Joint Hypothesis Testing

Joint Hypothesis Testing

$Y - b_0 + b_1 X_1 + b_2 X_2 \cdots$

- An F-test is used to test whether at least one slope coefficient is significantly different from zero.

H_0: $b_1 = b_2 = b_3 = \ldots = b_k = 0$; H_a: at least one $b_j \neq 0$ ($j = 1$ to k)

F-Statistic

$$\boxed{F = \frac{ESS/k}{SSR/(n-k-1)}}$$

$$F = \frac{S_1^2}{S_2^2}$$

- The F-test here is always a one-tailed test.
- The test assesses the effectiveness of the model as a whole in explaining the dependent variable.
- Decision rule: reject H_0, if $F_{(test\text{-}statistic)} > F_{c(critical\ value)}$.

F-Statistic

- When there is only **one regressor**, the F-statistic tests a single restriction, and the F-statistic is the square of the t-statistic.
- When there are **two regressors**, the joint null hypothesis has the two restrictions that $\beta_1 = 0, and\ \beta_2 = 0$, the F-statistic combines the two t-statistics t_1 and t_2 using the formula

$$F = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t_2}{1 - \hat{\rho}_{t_1,t_2}^2}\right)$$

  - Where $\hat{\rho}_{t_1,t_2}$ is an estimator of the correlation between the two t-statistics.
- When there are **more than two regressors**, this formula is <u>incorporated into regression software</u>, making the F-statistic easy to compute in practice.

F-Statistic

$t^2 = F$

$\frac{\hat{b}_1}{SE}$ $\frac{SSR/k}{SSE/(n-k-1)}$

- When there is only **one regressor**, the F-statistic tests a single restriction, and the F-statistic is the square of the t-statistic.
- When there are **two regressors**, the joint null hypothesis has the two restrictions that $\beta_1 = 0, and\ \beta_2 = 0$, the F-statistic combines the two t-statistics t_1 and t_2 using the formula

$$F = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t_2}{1 - \hat{\rho}_{t_1,t_2}^2}\right)$$

  - Where $\hat{\rho}_{t_1,t_2}$ is an estimator of the correlation between the two t-statistics.

- When there are **more than two regressors**, this formula is incorporated into regression software, making the F-statistic easy to compute in practice.

Homoskedasticity and Heteroskedasticity

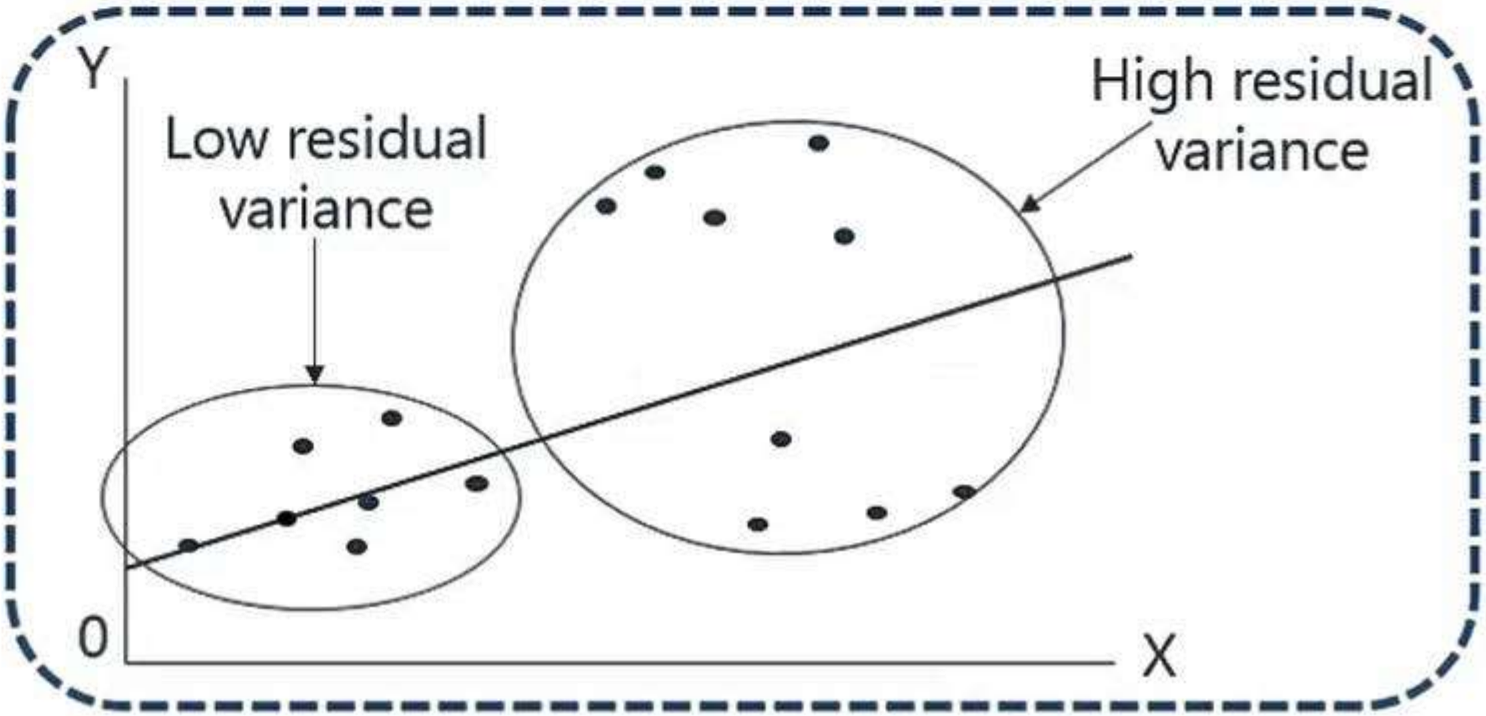
五、违反假设的后果

Homoskedasticity and Heteroskedasticity 异方差

- The error term ε_i is homoskedasticity if the variance of the conditional distribution of ε_i given X_i is **constant** for i = 1, …, n and in particular does not depend on X_i.

$V(\varepsilon_i) \neq$ constant

- Otherwise the error term is **heteroskedastic**.

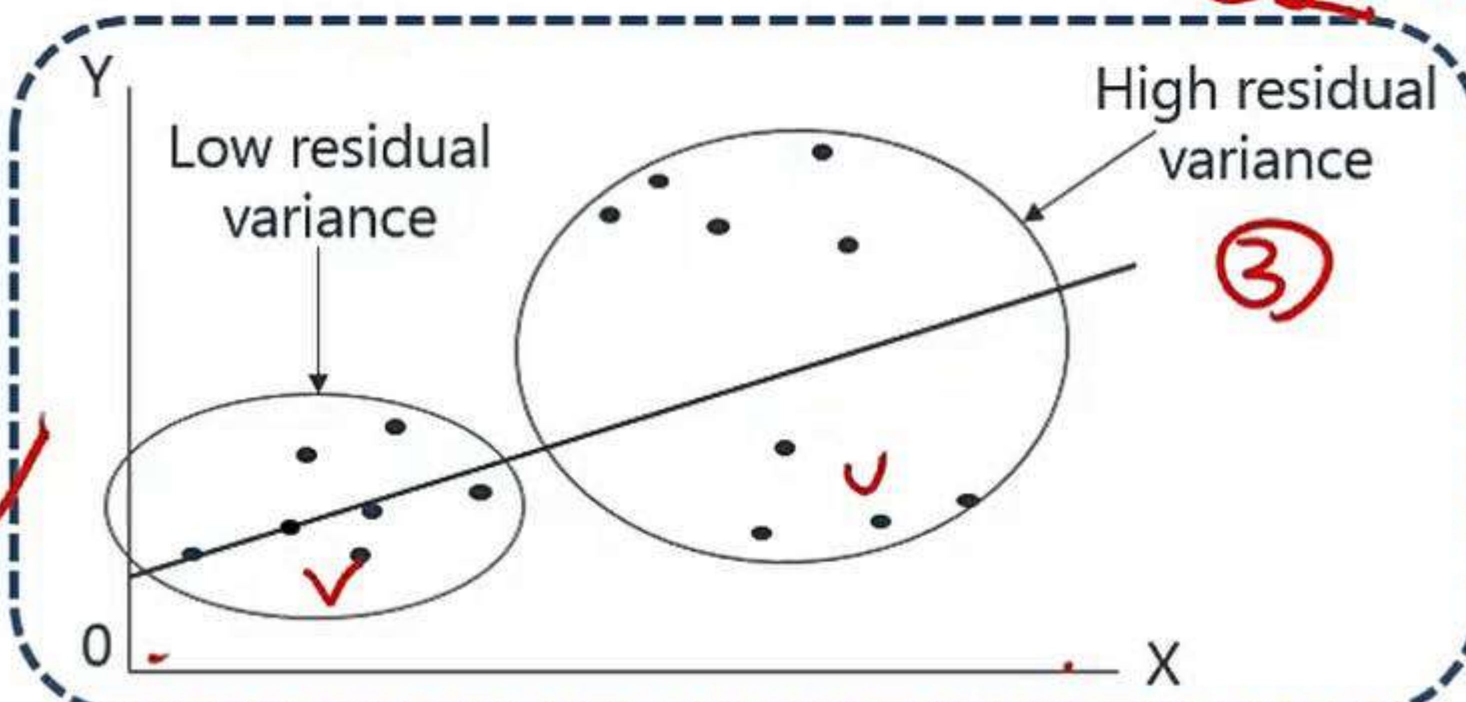


Homoskedasticity and Heteroskedasticity

五 违反假设的后果

Homoskedasticity and Heteroskedasticity 异方差

- The error term ε_i is homoskedasticity if the variance of the conditional distribution of ε_i given X_i is **constant** for i = 1, ..., n and in particular does not depend on X_i. ① $V(\varepsilon_i) \neq$ constant
- Otherwise the error term is **heteroskedastic**. ② $V(\varepsilon_1) =$ constant

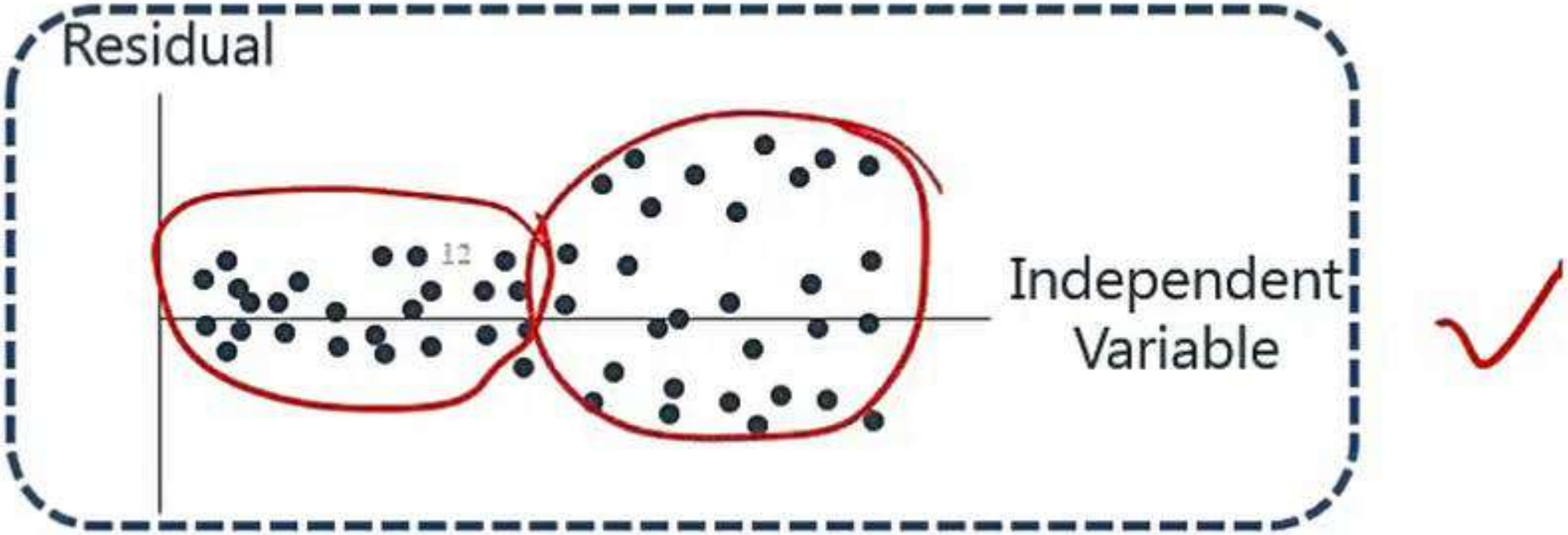


③

Homoskedasticity and Heteroskedasticity

Detecting Heteroskedasticity

- As shown in the figure below, the residuals and the scatter plot of the independent variables can show the relationship between the observations.



- The residual graph in the graph illustrates the existence of the conditional heteroscedasticity. Note that with the increase of independent variables, regression residuals how incremental changes.

Homoskedasticity and Heteroskedasticity

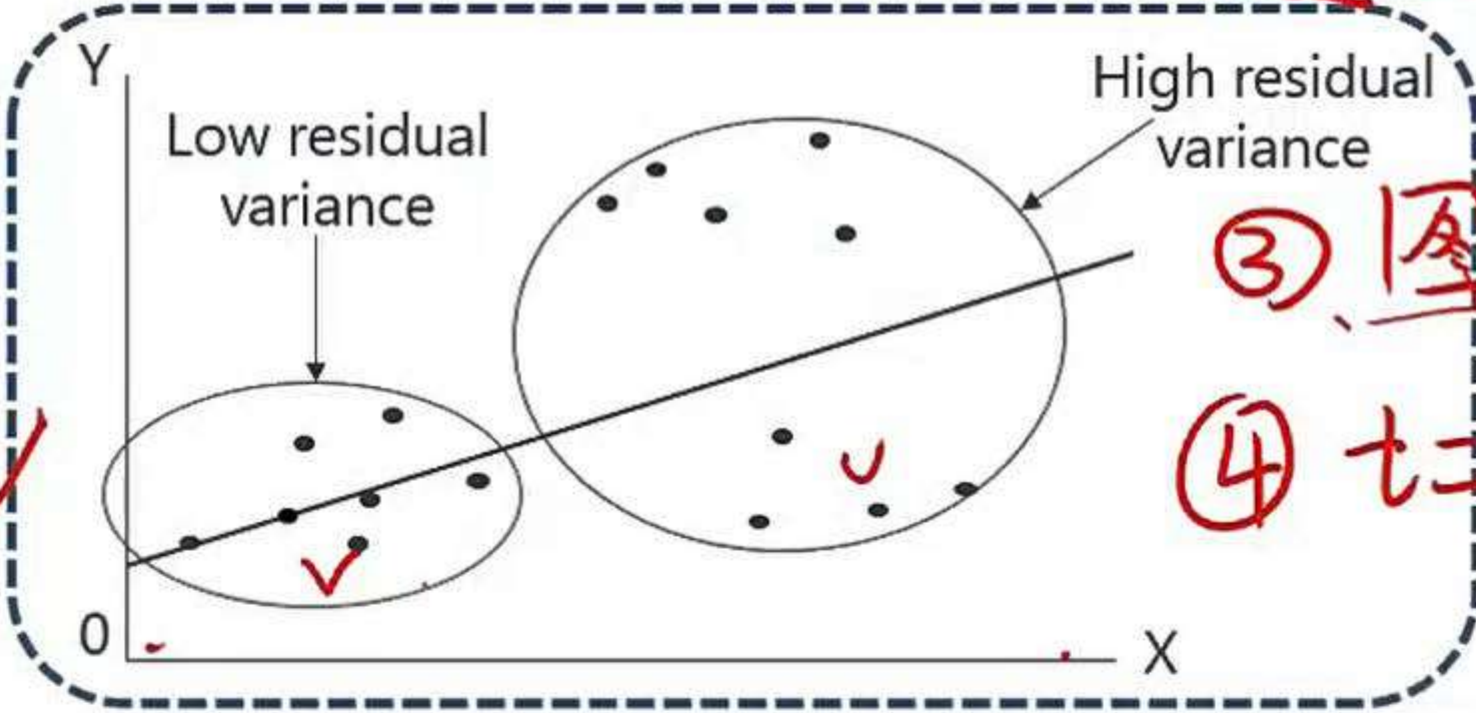
三 违反假设的后果.

Homoskedasticity and Heteroskedasticity 异方差

- The error term ε_i is homoskedasticity if the variance of the conditional distribution of ε_i given X_i is **constant** for i = 1, ..., n and in particular does not depend on X_i.
- Otherwise the error term is **heteroskedastic**.

① $V(\varepsilon_i) \neq$ constant

② $V(\varepsilon_i) =$ constant



③ 图

④ $t = \dfrac{b_1 - 0}{SE}$

b_1 → 不影响

SE ← 影响

Homoskedasticity and Heteroskedasticity

- **Effect of Heteroskedaticity on Regression Analysis**
  - The standard errors are usually not reliable estimate.
  - The coefficient estimates (the b_1) aren't influenced.
  - If the standard error is too small, but the coefficient estimates itself is not affected, the t-statistic will become too large, no statistically significant null hypothesis is rejected too often.
    - ✓ The opposite will be true if the standard errors are too large.

SE

Serial Correlation (autocorrelation)

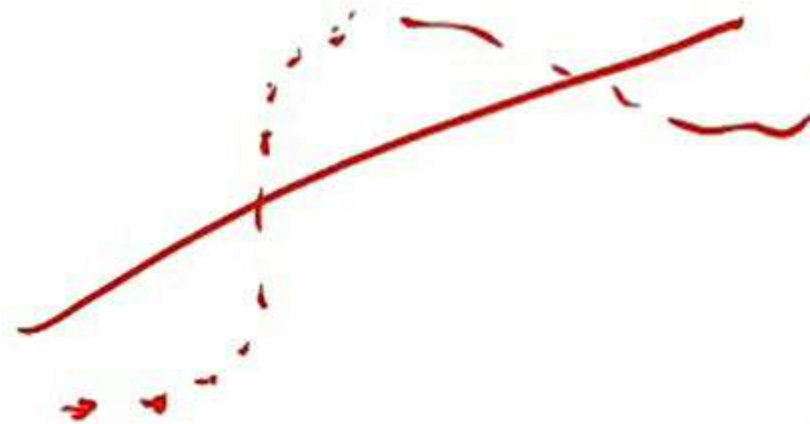
②.Serial correlation (autocorrelation) 自相关

- Serial correlation (autocorrelation) refers to the situation that the error terms are correlated with one another.
- Serial correlation is often found in time series data.
- Positive serial correlation exists when a positive regression error in one time period increases the probability of observing regression error for the next time period.
- Negative serial correlation occurs when a positive error in one period increases the probability of observing a negative error in the next period.

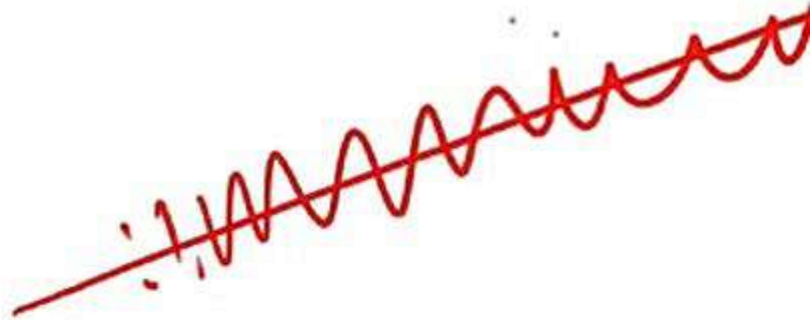
① 定义：ε_t 与 ε_{t-1} 是相关. ② 违反. uncorrelated

③ 分类. { 正 / 负

3. 分类 {
正



负



4 影响：$\underline{t} = \frac{\hat{b}_1}{SE}$

Multicollinearity

线性相关

3 ➢ **Multicollinearity** 多重共线性：X_1 与 X_2 之间高度

- Multicollinearity refers to the case where two or more independent variables are highly interrelated..
- In practice, multicollinearity is often a matter of degree rather than of absence or presence.

➢ **Two methods to detect multicollinearity**

- t-test indicate that none of the individual coefficients is significantly different than zero, while the F-test indicates overall significance and the R^2 is high.
- The absolute value of the correlation between any two sample independent variable is greater than 0.7(i.e., $|r| > 0.7$).

➢ **Methods to correct multicollinearity**

- Omit one or more of the correlated independent variables.

$$y_t = b_0 + b_1 \varepsilon_t + b_2 \varepsilon_{t-1}$$

自相关

$$y_t = b_0 + b_1 X_1 + b_2 X_2$$

工资 学年 X_1 经 X_2 相关

Multicollinearity

线性相关

3 ➢ **Multicollinearity** 多重共线性：X_1 与 X_2 之间高度

- Multicollinearity refers to the case where two or more independent variables are highly interrelated..
- In practice, multicollinearity is often a matter of degree rather than of absence or presence. 判断多重共线性方法

② ➢ **Two methods to detect multicollinearity** (1) common sense.

- (2) t-test indicate that none of the individual coefficients is significantly different than zero, while the F-test indicates overall significance and the R^2 is high. F可以通过，t不能通过
- (3) The absolute value of the correlation between any two sample independent variable is greater than 0.7(i.e., $|r| > 0.7$). X_1, X_2.

$|\rho| > 0.7$

➢ **Methods to correct multicollinearity**

- Omit one or more of the correlated independent variables.

③ 解决方式、

Omitted Variable Bias

4 Omitted Variable Bias 重大遗漏变量

- Omitted variable bias is the bias in the OLS estimator that arises when one or more included regressors are <u>correlated with an omitted variable</u>.
- For omitted variable bias to arise, two things must be true:
  - ✓ At least one of the included regressors must be correlated with the omitted variable.
  - ✓ Omitted variables must be the determinant of the dependent variable Y.

X_3

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \varepsilon$$

- According to the datas in the table ,and calculate the F-statistic is closed to:

A. 46.61
B. 67.16
C. 64.84
D. 54.03

$$\frac{2000}{29.78}$$

	df	SS	MSS
Regression	3	ESS	2000
Residual	46	SSR	29.78
Total	49	TSS	-

- Correct Answer : B

➢ A factor analysis of the dividend-adjusted returns of ABC Ltd.'s stock price was undertaken to determine which economic factors contributed to its performance. The regression was performed on 460 observations. The results are as follows:

Predictor	Coefficient	Standard Error of Coefficient
Intercept	-0.0243	0.005772
All share index	0.0256	0.017655
Industrial index	0.0469	0.006398
Financial index	0.0012	0.001412

Regression sum of Squared (RSS)	12,466.47
Sum of Squared Errors (SSE)	1,013.22
Sum of Squared Total (SST)	13,479.69

Which one of the following options correctly describes which variables are significant at the 5% level, and the R^2 statistic, respectively?

	Significant Variables at 5% level	R^2 statistic
A.	Intercept; Industrial index	0.924834
B.	Intercept; Industrial index	0.075166
C.	All share index; Industrial Index	0.924834
D.	All share index; Industrial Index	0.075166

A

- A factor analysis of the dividend-adjusted returns of ABC Ltd.'s stock price was undertaken to determine which economic factors contributed to its performance. The regression was performed on 460 observations. The results are as follows:

$n=460$

Predictor	Coefficient	Standard Error of Coefficient
Intercept	-0.0243	0.005772
All share index	0.0256	0.017655
Industrial index	0.0469	0.006398
Financial index	0.0012	0.001412

Regression sum of Squared (RSS)	12,466.47
Sum of Squared Errors (SSE)	1,013.22
Sum of Squared Total (SST)	13,479.69

Which one of the following options correctly describes which variables are significant at the 5% level, and the R^2 statistic, respectively?

	Significant Variables at 5% level	R^2 statistic
A.	Intercept; Industrial index	0.924834
B.	Intercept; Industrial index	0.075166
C.	All share index; Industrial Index	0.924834
D.	All share index; Industrial Index	0.075166

$\frac{ESS}{TSS}$

- Correct Answer: A
  - The following table shows the test statistics for each of the four variables, calculated by dividing the variable coefficient by the standard error. The variable is significant if the absolute value of the t-test is greater than the critical value from the student's t-distribution for 456 degrees of freedom (which is very close to the z-statistic since the number of observations is so high), i.e. 1.96.

Predictor	T-stat	Significant
Intercept	-4.21	Yes
All share index	1.45	No
Industrial index	7.33	Yes
Financial index	0.85	No

Example 2

➢ A regression of a stock's return (in percent) on an industry index's return (in percent) provides the following results:

	Coefficient	Standard Error
Intercept	2.1	2.01
Industry index	1.9	0.31
	Degrees of Freedom	**SS**
Explained	1	92.648
Residual	3	24.512
Total	4	117.160

Which of the following statements regarding the regression is incorrect?

A. The correlation coefficient between the X and Y variables is 0.889.

B. The industry index coefficient is significant at the 99% confidence interval.

C. If the return on the industry index is 4%, the stock's expected return is 9.7%.

D. The variability of industry returns explains 21% of the variation of company returns.

Example 2

$\boxed{Y=2.1+1.9X}$

$\frac{1.9}{0.31}=6.13$

D

- A regression of a stock's return (in percent) on an industry index's return (in percent) provides the following results:

t-

$\frac{2.1}{2.01}$

	Coefficient	Standard Error
Intercept	2.1	2.01
Industry index	1.9	0.31
	Degrees of Freedom	**SS**
Explained	1	92.648
Residual	3	24.512
Total	4	117.160

$\frac{92.648}{117.16}=R^2$

$R^2=\rho^2$

Which of the following statements regarding the regression is incorrect?

A. The correlation coefficient between the X and Y variables is 0.889.

$\rho=0.889$

B. The industry index coefficient is significant at the 99% confidence interval.

C. If the return on the industry index is 4%, the stock's expected return is 9.7%.

D. The variability of industry returns explains 21% of the variation of company returns.

R^2

总结：

一. Term：$Y = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n + \varepsilon$

二. Assumptions：6个

三. ANUVA Table：

1. 衡量线性回归拟合程度 R^2 Adjust R^2

2. $R^2 = \frac{ESS}{TSS}$

Ad. $R^2 = 1 - \frac{SSR/n-k-1}{TSS/n-1}$

$= 1 - \frac{n-1}{n-k-1}(1-R^2)$

3. 了解表格中所有数据来源

四 假设检验 1. t-检验：检验单个自变量对于Y的解释力度.

$$H_0: b_1 = 0$$

$$t = \frac{\hat{b}_1}{S_{\hat{b}_1}}$$

2. F检验：检验所有自变量对于Y共同的解释力度.

$$H_0: b_1 = b_2 = b_3 \cdots = 0$$

$$F = \frac{RSS/k}{SSR/(n-k-1)}$$

五. 违反假设检验情况：

1. 异方差 ① 定义：$V(\varepsilon) \neq$ constant

② 图：2个

③ 影响：$t = \dfrac{\hat{b}_1}{SE}$ → 不影响 $\hat{b}_1$；SE → 影响

2. 自相关 ① 定义：ε_t 与 ε_{t-1} 有关

② 图/分类：

正　　负

③ 影响 SE，不影响 $\hat{b}_1$

3. 多重共线性：①定义：X_1 与 X_2 高度线性相关

② 判断方法 (1) common sense

(2) F通过，t不通过

(3) $|\rho_{x_1,x_2}| > 0.7$

③ 解决方法：omit one variable 重新做回归

④ 影响：b_i 影响

4. 重大遗漏变量：与多重共线性反向的过程

eg: GDP.

Forecasting Trends

时间序列分成三个模型

一、Trend：分类、评价模型的好坏的标准

二、cycle、来源、怎么描述

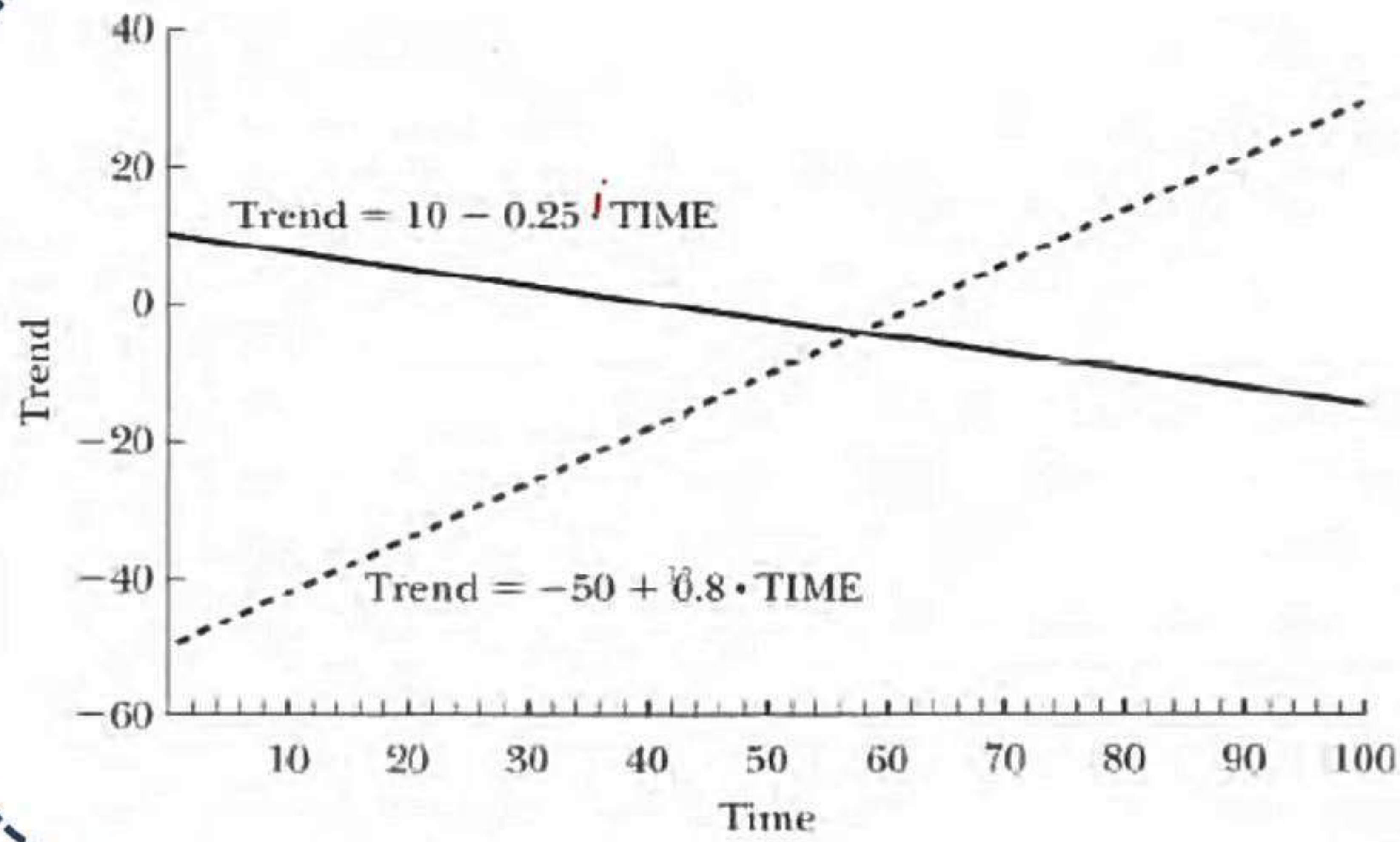
三、Noise.

1.协方差平稳
2.白噪声
3.滞后因子
4 Wold理论
5 AR、MA、ARMA

Modeling and Forecasting Trend

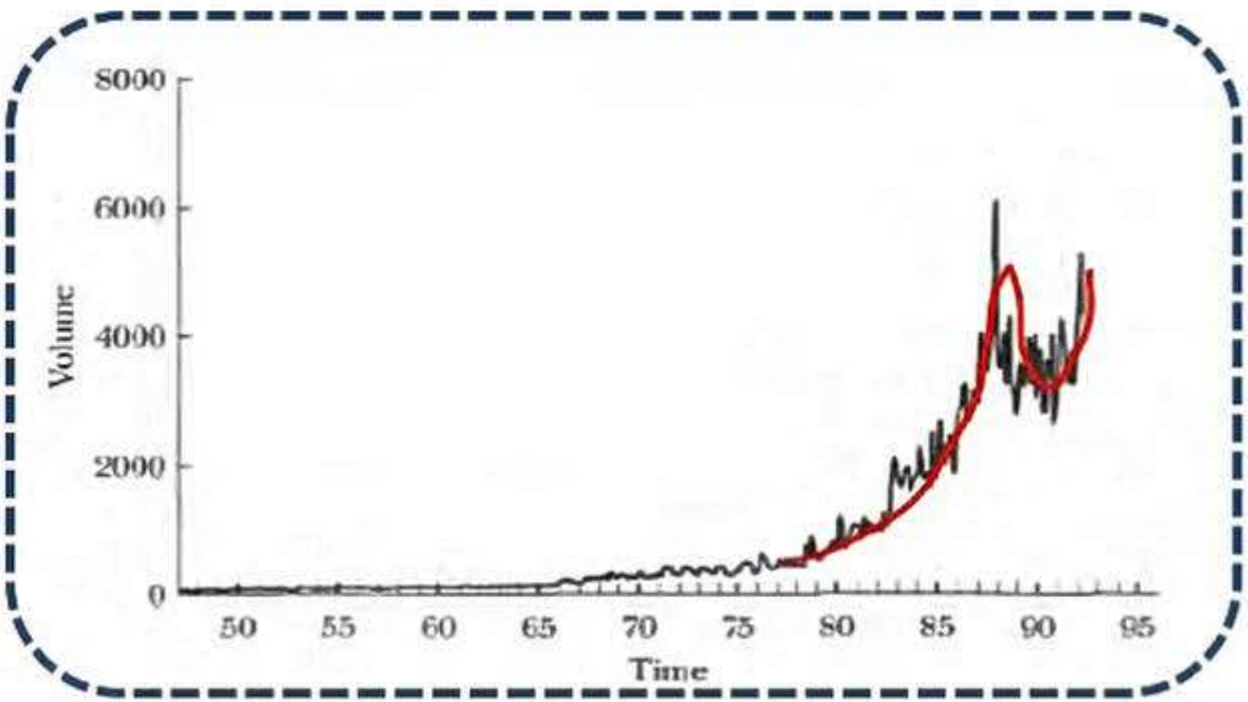
1、趋势分类

➢ **Linear trend**: The trend in which appears roughly linear, meaning that it increases or decreases like **a straight line**.



Modeling and Forecasting Trend

- **Non-linear Trend**
- Sometimes trend appears **nonlinear**, or curved, as, for example, when a variable increases **at an increasing or decreasing rate**. Ultimately, we don't require that trends be linear, only that they be smooth. Next figure shows the monthly volume of shares traded on the New York Stock Exchange (NYSE). Volume increases at **an increasing rate**; the trend is therefore **nonlinear**.



Modeling and Forecasting Trend

- **Quadratic trend models** can potentially capture nonlinearities such as those observed in the volume series. Such trends are quadratic, as opposed to linear, functions of time,

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t{}^2$$

Y = ax² + bx + c

- **Polynomial trend:**

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t{}^2 + \beta_3 TIME_t{}^3 + \ldots + \beta_n TIME_t{}^n$$

- **Exponential trend, or log-linear trend**, and is very common in business, finance, and economics. That's because economic variables often display roughly **constant growth rates** (for example, 3% per year). If trend is characterized by constant growth at rate β_1 , then we can write

$$T_t = \beta_0 e^{\beta_1 TIME_t}$$

Measure of Model fitness

评价模型的好坏标准.

- There are four ways to measure the fitness of the model
  - MSE
  - S^2
  - AIC
  - SIC

Measure of Model fitness

评价模型的好坏标准

- There are four ways to measure the fitness of the model
  - MSE
  - S^2
  - AIC
  - SIC

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

数值越小越好

Modeling and Forecasting Trend

- Measure 1: Mean Squared Error (MSE) .
  - The model with least MSE would be chosen for fitting the data series.

$$\boxed{MSE = \frac{\sum_{t=1}^{T} e_t^2}{T}}$$

- Measure 2: S^2.
  - As degree of freedom represents the choice of freely selecting the variables during the model fitting exercise therefore, to reduce the MSE bias, the degree of freedom must be deducted from the sample size to arrive at adjusted MSE, commonly referred as S^2.

$$\boxed{S^2 = \frac{\sum_{t=1}^{T} e_t^2}{T - k}}$$

Modeling and Forecasting Trend

- Measure 3: The Akaike information criterion (AIC). 渐近有效
  - AIC offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. In doing so, **it deals with the trade-off between the goodness of fit of the model and the complexity of the model.**

$$AIC = e^{\left(\frac{2k}{T}\right)} \boxed{\frac{\sum_{t=1}^{T} e_t^2}{T}} \longrightarrow \text{MSE}$$

帮助最快选择想要in模型

Modeling and Forecasting Trend

- Measure 4: The **Schwarz information criterion(SIC).**
  - The SIC is an increasing function of unexplained variation in the dependent variable and the number of explanatory variables. **Hence, lower SIC implies either fewer explanatory variables, better fit, or both**.

$$\boxed{SIC = T^{\left(\frac{k}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T}}$$

Measure of Model fitness

评价模型的好坏标准.

- There are four ways to measure the fitness of the model

说:

- MSE：未考虑自由度.
- S^2
- AIC：渐近有效.
- SIC：平稳的 一致性.

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

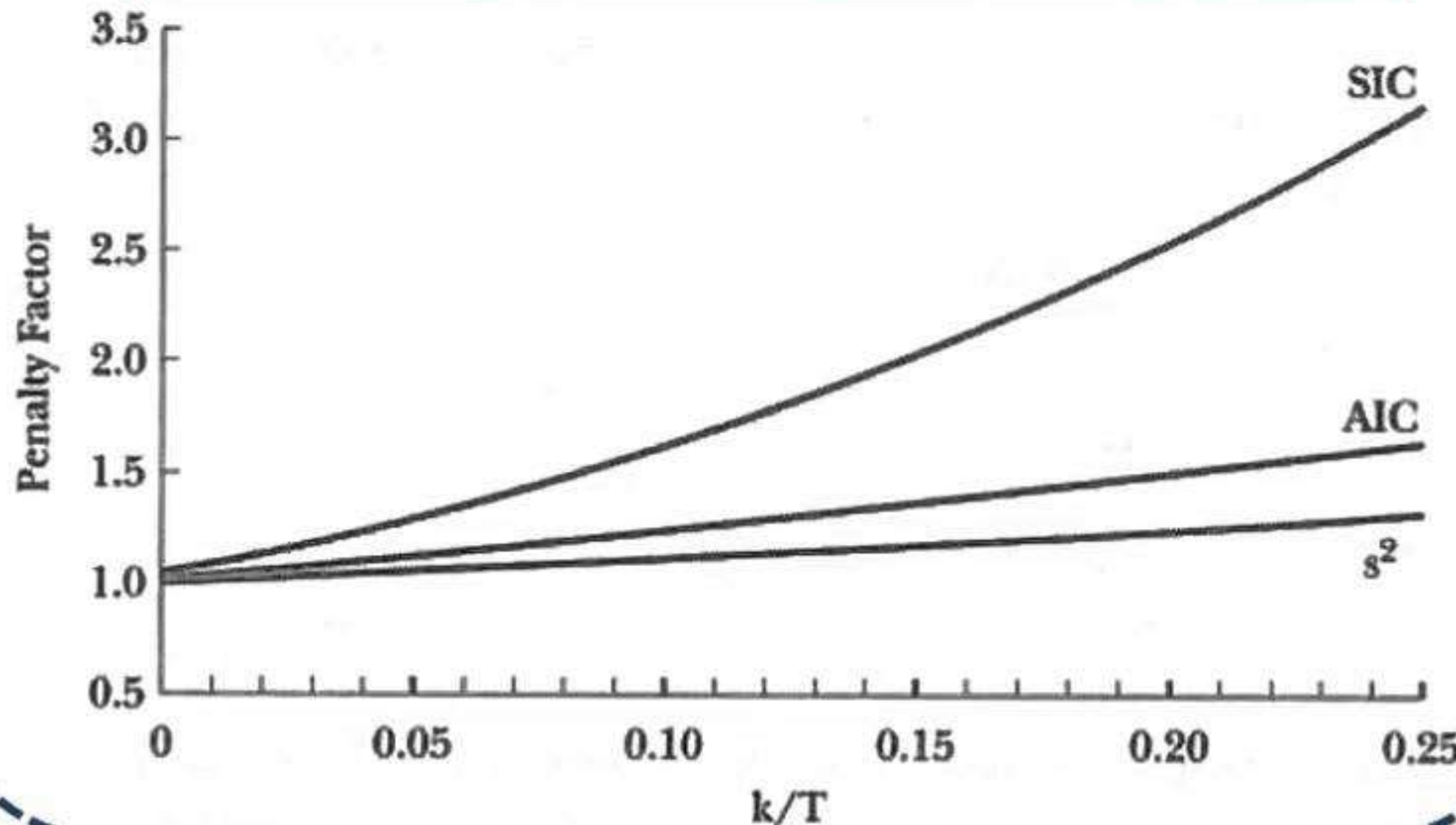
数值越小越好.

Modeling and Forecasting Trend

n↑ 误差 ↓

➤ The SIC (the only **consistency** ceritria) generally penalizes free parameters more strongly than does the Akaike information criterion, though it depends on the size of T and relative magnitude of T and k.

Degrees-of-freedom penalties, various model selection criteria.

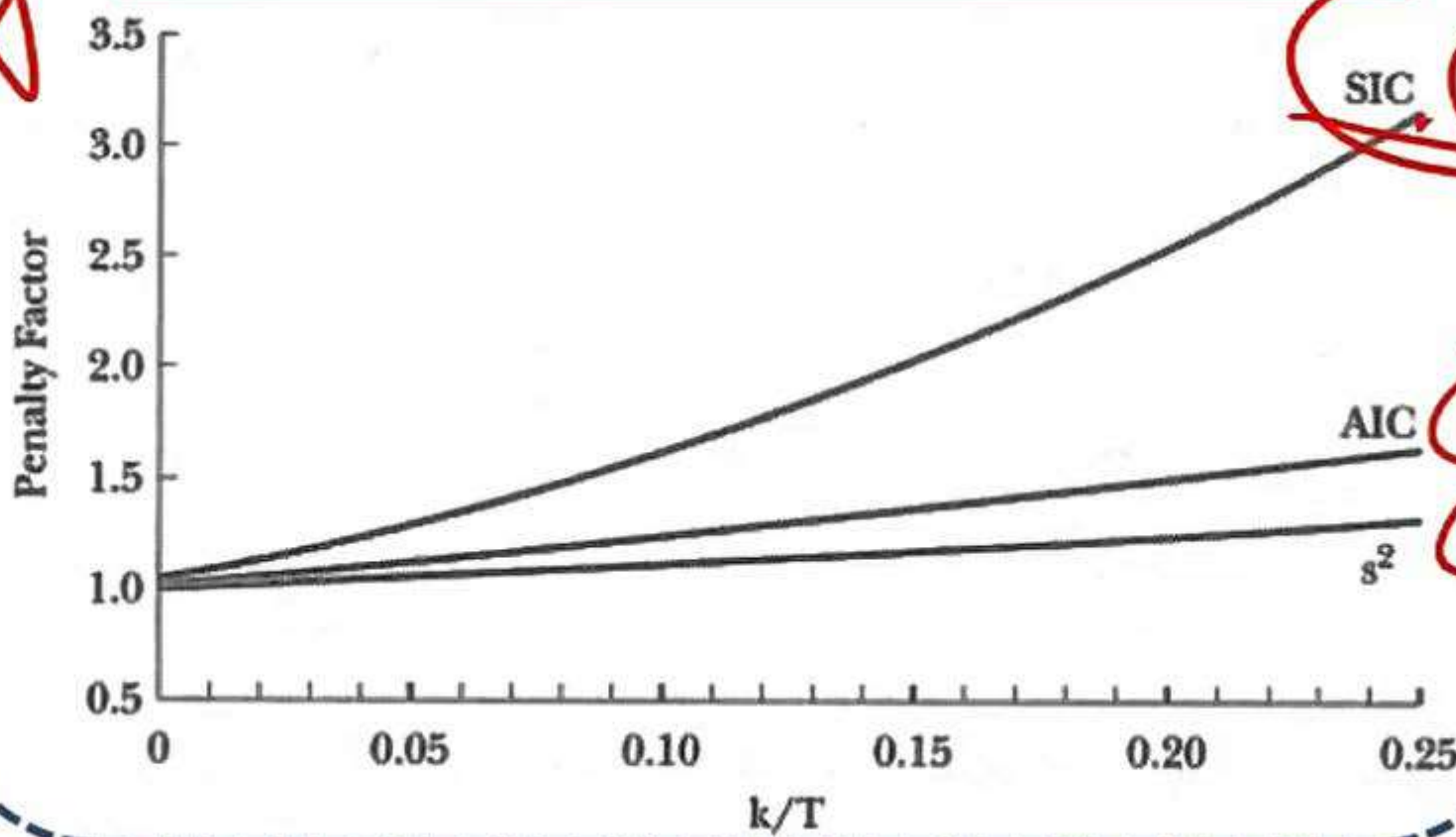


Modeling and Forecasting Trend

n↑ 误差↓

- The SIC (the only **consistency** ceritria) generally penalizes fr parameters more strongly than does the Akaike information criterio though it depends on the size of T and relative magnitude of T and k.

Degrees-of-freedom penalties, various model selection criteria.



Modeling and Forecasting Trend

- We evaluate model selection criteria in terms of a key property called **consistency.** A model selection criterion is consistent if the following conditions are met:
  - as the sample size gets large, the chosen measure will choose the true model correctly or with the biggest probability.
  - SIC is consistency while others are not.
- If the no model discussed has the property of consistency. We're then led to a different optimality property, called **asymptotic efficiency.**
  - as the sample size gets large, an asymptotically efficient model selection criterion chooses the model has the fastest speed to approach to the true error variance.
  - The AIC, although inconsistent, is asymptotically efficient, whereas the SIC is not.

渐近有效：

苹果　　A 梨　　B 香蕉　　C 芒果

AIC

n=50　AIC　　SIC

⋮

n=10000　SIC

GOLDEN FUTURE

Modeling and Forecasting Seasonality

周期 cycle

- **The Sources Of Seasonality**
  - Any **technology** that involves the weather, such as production of agricultural commodities, is likely to be seasonal as well.
  - **Preferences** may also be linked to the calendar. People want to do more vacation travel in the summer, which tends to increase both the price and quantity of summertime gasoline sales.
  - **Social institutions** that are linked to the calendar, such as holidays.
- **A key technique for modeling seasonality is regression on seasonal dummies.**

Trend
cycle
noise

Trend
seasonality
cycle

Modeling and Forecasting Seasonality

周期 cycle

- **The Sources Of Seasonality** 1、来源.
  - ① Any **technology** that involves the weather, such as production of agricultural commodities, is likely to be seasonal as well. 大棚技术
  - ② **Preferences** may also be linked to the calendar. People want to do more vacation travel in the summer, which tends to increase both the price and quantity of summertime gasoline sales. 偏好
  - ③ **Social institutions** that are linked to the calendar, such as holidays.
- **A key technique for modeling seasonality is regression on seasonal dummies.**

Trend
cycle
noise

Trend.
seasonality
cycle

Modeling and Forecasting Seasonality

- The pure seasonal dummy model is 哑变量

$$y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

- Trend may be included as well, in which case the model is

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

- Expand the seasonality model: calendar effects.
  - **Holiday variation** refers to the fact that some holidays' dates change over time.
  - **Trading-day variation** refers to the fact that different months contain different numbers of trading days or business days

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \sum_{i=1}^{v1} \delta_i^{HD} HDV_{it} + \sum_{i=1}^{v2} \delta_i^{TD} TDV_{it} + \varepsilon_t.$$

$$\boxed{Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + \varepsilon}$$ ✓

X_1 0 不是第一季度.

1：第一季度产量. $b_1 + b_0$

X_2 0

1 二

X_3 0

1 三

b_0 第四季度

Modeling cycle

- **What is cycle?** noise
  - As we mentioned the cycles, we generally considered the all-encompassing notion of cyclicality into two parts:
    - ✓ Can be captured by trend or seasonal
    - ✓ Cannot be captured by trend or seasonal
- **Charactering cycle**
- **Modeling cycle**

Characterizing Cycles

- **Why we need covariance stationary?** 协方差平稳
  - If the underlying probabilistic structure of the series were changing over time, there would be no way to predict the future accurately on the basis of the past, **because the laws governing the future would differ from those governing the past**.
  - 定 If we want to forecast a series, at minimum we require the mean and covariance to be stable and finite over time, which we call it **covariance stationary**.
  - In this chapter all we mentioned covariance stationarity is called **second-order stationarity** or **weak stationarity** .It means that a series whose mean and variance and covariance are stable and finite all the time but the skewness and kurtosis are not necessary.

Characterizing Cycles

2 ➢ **White noise**:

- In general, if there is a process that have zero mean, constant variance, and no serial correlation, the process is called **zero-mean white noise, or simply white noise**. Sometimes for short we write:

$$\varepsilon_t \sim WN(0,\ \sigma^2)$$

$$y_t = \varepsilon_t$$

hence:

$$y_t \sim WN(0,\ \sigma^2)$$

Characterizing Cycles

2 ➤ **White noise** ①定义 ② 分类

- In general, if there is a process that have zero mean, constant variance, and no serial correlation, the process is called **zero-mean white noise, or simply white noise**. Sometimes for short we write:

(1) (2) (3)

weak white noise.

$$\varepsilon_t \sim WN(0,\ \sigma^2)$$

$$y_t = \varepsilon_t$$

hence:

$$y_t \sim WN(0,\ \sigma^2)$$

Characterizing Cycles

- **Independent white noise** ① $\mu=0$ σ^2 = constant
  - There is a point to be mentioned that ε_t and y_t are serially uncorrelated, they are not necessarily normally distributed. If y is serially independent, then we say that y is **independent white noise**. We write:

$$y_t \overset{iid}{\sim} (0, \sigma^2)$$

Strong white noise

  - Another name for independent white noise is **strong white noise,** in contrast to standard serially uncorrelated, **weak white noise**.
- **Normal white noise** $\mu=0$ σ^2 = constant
  - If y is serially uncorrelated and normal distributed, of course, y also is a serially independent, we will say that y is **normal white noise** or **Gaussion white noise.**

Characterizing Cycles

3 Hypothesis test for white noise

- Hypothesis testing for one autocorrelation coefficient is not enough to conclude covariance stationary, instead, **a joint hypothesis test for all the autocorrelation coefficients to be zero is needed.**
- **Box-Pierce Q-Statistic & Ljung-Box Q-Statistic**

Characterizing Cycles

3 Hypothesis test for white noise

- Hypothesis testing for one autocorrelation coefficient is not enough to conclude covariance stationary, instead, **a joint hypothesis test for all the autocorrelation coefficients to be zero is needed.**
- **Box-Pierce Q-Statistic & Ljung-Box Q-Statistic**

$\mu=0$ $Var=$ constant $\rho=0$

t, z χ^2

Characterizing Cycles

$\rho(1)$

Box-Pierce Q-Statistic & Ljung-Box Q-Statistic

- H_0: The series are white noise and hence autocorrelation of the series is zero.
- For Box-Pierce Q-statistic, the formula used is:

$\rho(\tau)$

$$\boxed{Q_{BP} = T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau)}$$

- Whereas in case of Ljung-Box Q-statistic, the test statistic is derived as:

$$\boxed{Q_{LP} = T(T+2)\sum_{\tau=1}^{m}\left(\frac{1}{T-\tau}\right)\hat{\rho}^2(\tau)}$$

  - ✓ T = Sample size
  - ✓ m = the maximum lag under observation
- Reject the null when Q-statistic is large.(look up in the χ^2 table)

X	X_1	X_2	X_3	X_4	X_5.
Y	X_2	X_3	X_4	X_5	

$\rho(1)$ 滞后一期

X	X_1	X_2	X_3	X_4	X_5
	X_2	X_3	X_4	X_5	
	X_3	X_4	X_5		

$\rho(1)$ 滞后一期 r

$\rho(2)$ 二

Characterizing Cycles

$\rho(1)$

- **Box-Pierce Q-Statistic & Ljung-Box Q-Statistic**
  - H_0: The series are white noise and hence autocorrelation of the series is zero.

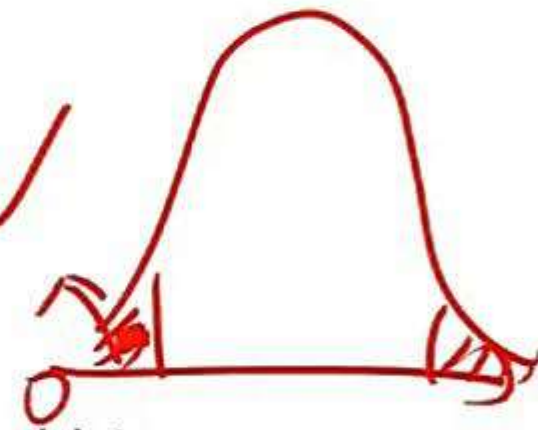
    $H_0: \rho = 0$
  - For Box-Pierce Q-statistic, the formula used is:

    $\rho(\tau)$

$$\boxed{Q_{BP} = T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau)}$$

  - Whereas in case of Ljung-Box Q-statistic, the test statistic is derived as:

$$\boxed{Q_{LP} = T(T+2)\sum_{\tau=1}^{m}\left(\frac{1}{T-\tau}\right)\hat{\rho}^2(\tau)}$$



    ✓ T = Sample size

    ✓ m = the maximum lag under observation
  - Reject the null when Q-statistic is large.(look up in the χ^2 table)

Characterizing Cycles

3. ➤ **Lag operators:** 滞后项 P(1)

- The lag operator can be explained very simply that it just "operates" on a series by lagging it, just described it as follow:

$$Ly_t = y_{t-1}$$

Similarly,

$$L^2 y_t = L(L(y_t)) = L(y_{t-1}) = y_{t-2}$$

Characterizing Cycles

- We 'll also operate on a series not with the lag operator but with a **polynomial in the lag operator.** A lag operator polynomial of degree *m* is just a linear function of powers of L, up through the m^{th} power, like this:

$$B(L) = b_0 + b_1 L + b_2 L^2 + \ldots\ldots + b_m L^m$$

Characterizing Cycles

4 **Wold's theorem(Wold's representation)**

- Let $\{y_t\}$ be any zero-mean covariance-stationary process.

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \qquad \varepsilon_t \sim WN(0, \sigma^2)$$

✓ where the b_i are coefficients with $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$

✓ In short, the correct "model" for any covariance stationary series is some infinite distributed lag of white noise, called the **Wold's representation.** The ε_t are often called **innovations.**

Characterizing Cycles

4 **Wold's theorem(Wold's representation)**

- Let $\{y_t\}$ be any zero-mean covariance-stationary process.

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \qquad \varepsilon_t \sim WN(0, \sigma^2)$$

✓ where the b_i are coefficients with $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$ 收敛

✓ In short, the correct "model" for any covariance stationary series is some infinite distributed lag of white noise, called the **Wold's representation.** The ε_t are often called **innovations.**

Characterizing Cycles

- **Rational distributed lags**
  - The Wold's representation points to the crucial importance of models with **infinite distributed lags**. But it is not suitable for practical cases, so we transfer the infinite polynomials to finite polynomials. Such polynomials are called rational polynomials, and distributed lags constructed from them are called **rational distributed lags**.
    - ✓ Where the numerator polynomial is of degree q ,

$$\boxed{\Theta(L) = \sum_{i=0}^{q} \theta_i L^i}$$

Modeling Cycle

必 **Describe the types of cyclical models**

- MA
- AR
- ARMA

估计建模 noise

Modeling Cycle

➤ **MA(1) Model** : Moving Average

$$\boxed{y_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1+\theta L)\varepsilon_t;\ \varepsilon_t \sim WN(0,\sigma^2)}$$

- The current value of the observed series is expressed as a function of current and lagged unobservable shocks.
- It is the very special case of Wold's representation.

➤ **MA(q) Model**

$$\boxed{\begin{gathered} y_t = \varepsilon_t + \theta\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} = \Theta(L)\varepsilon_t;\ \varepsilon_t \sim WN(0,\sigma^2) \\ \Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q \end{gathered}}$$

Modeling Cycle

- **AR(1) Model** : Auto correlation.

$$y_t = \varphi y_{t-1} + \varepsilon_t;\ \varepsilon_t \sim WN(0, \sigma^2)$$
$$(1 - \varphi L) y_t = \varepsilon_t$$

- **The relationship between AR and MA model**
  - The AR model described a relationship between y_t and y_{t-i}
  - The MA model described a relationship between y_t and ε_t ,which ε_t is a white noise process.

- AR(p) Model

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t;\ \varepsilon_t \sim WN(0, \sigma^2)$$
$$\Phi(L) y_t = \left(1 - \varphi_1 L - \varphi_2 L^2 + \ldots - \varphi_p L^p\right) y_t = \varepsilon_t$$

Modeling Cycle

ARMA(p, q) Model

- ARMA models are often both highly accurate and highly parsimonious.

$$y_t = \underline{\varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p}} + \theta \underline{\varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t}; \varepsilon_t \sim WN(0, \sigma^2)$$

- Regardless of autocorrelation or partial autocorrelation, their graphs all appear to be gradual damped.

Characterizing Cycles

工具 ①指标：autocorrelation 自相关

- The autocovariance is just the covariance between y_t and $y_{t-\tau}$, as the series is covariance stationary, so the autocorrelation will only depend on τ, and have no relationship with t , so the function can be written as follow:

partial autocorrelation 偏自相关

$$\gamma = (\tau) = \operatorname{cov}(y_t, y_{t-\tau}) = E(y_t - u)(y_{t-\tau} - u)$$

- The autocorrelations are just the "simple" or "regular" correlations between y_t and $y_{t-\tau}$, so the function can be written as follow:

②图.

$$\rho(\tau) = \frac{\operatorname{cov}(y_t, y_{t-\tau})}{\sqrt{\operatorname{var}(y_t)\sqrt{\operatorname{var}(y_{t-\tau})}}} = \frac{\gamma(\tau)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(\tau)}{\gamma(0)}$$

Characterizing Cycles

- The partial autocorrelations function measures the relationship between y_t and $y_{t-\tau}$ removed the effects of y_{t-1} ...... $y_{t-\tau+1}$, in other words, the partial autocorrelations function measures the relationship only between y_t and $y_{t-\tau}$.

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_n y_{t-n} + \varepsilon_t$$
$$p(\tau) = b_\tau$$

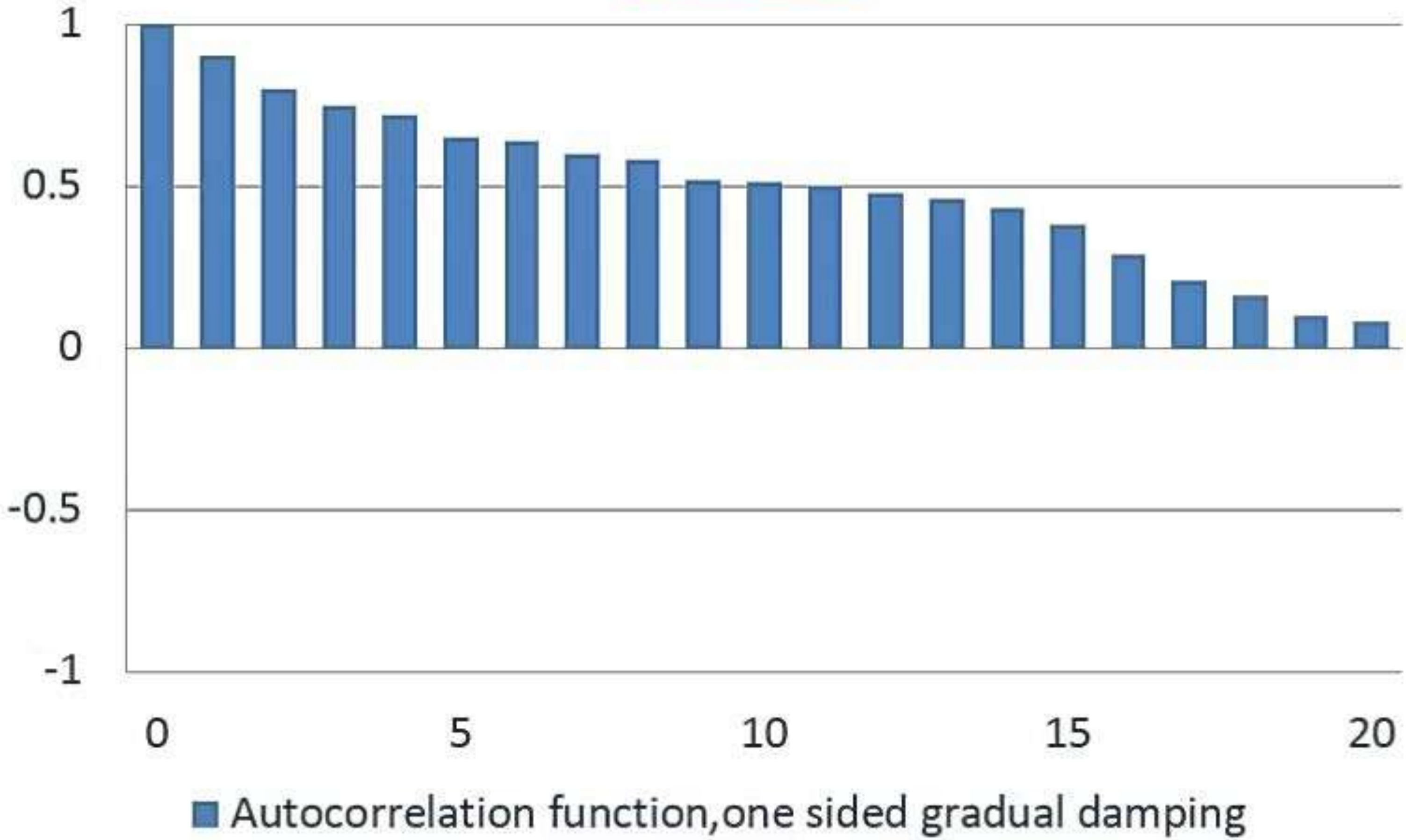
$$y_t = b_1 y_{t-1}$$

$$\rho$$

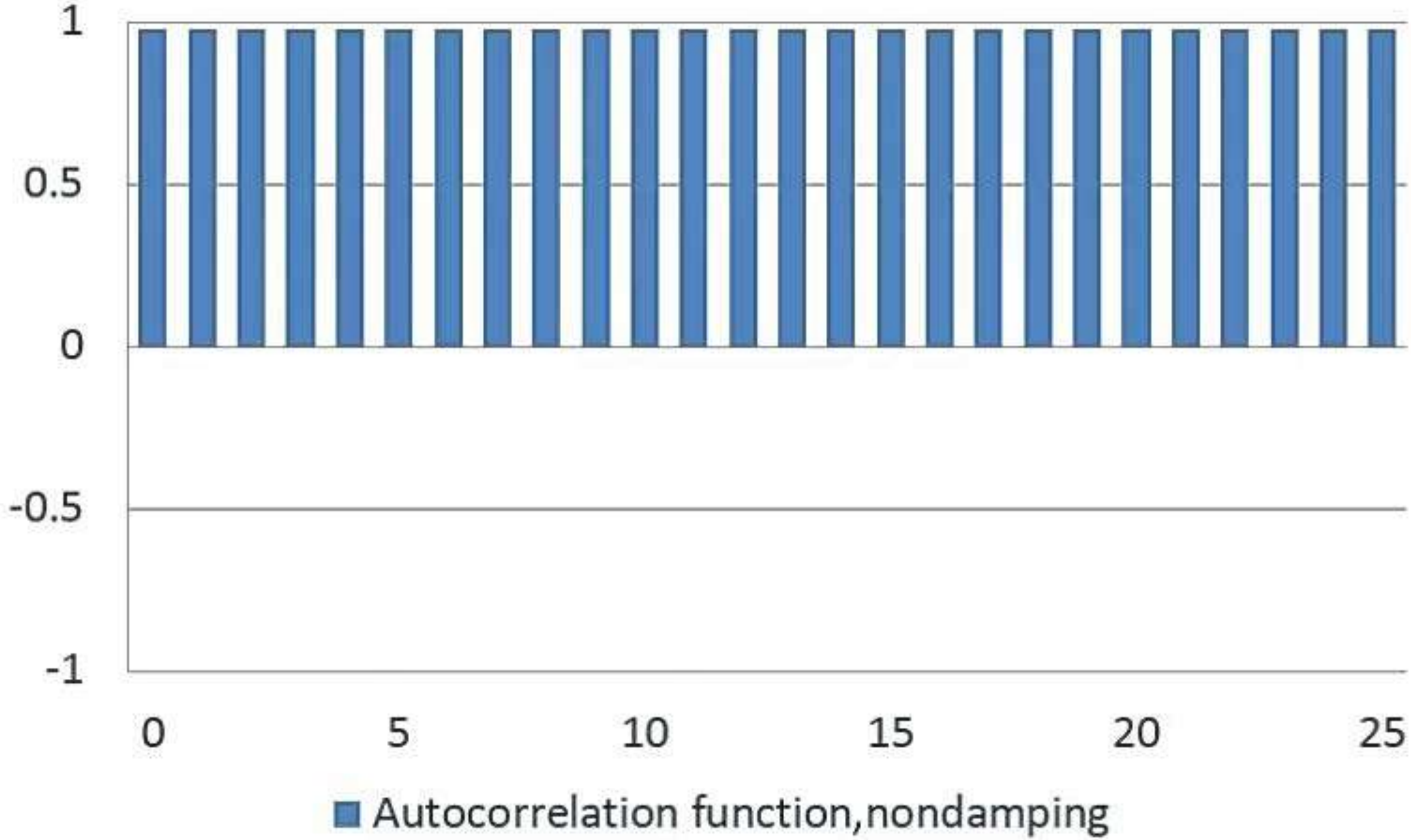
$$y_{t-1} = \cancel{b}\, a\, y_{t-2}$$

Characterizing Cycles

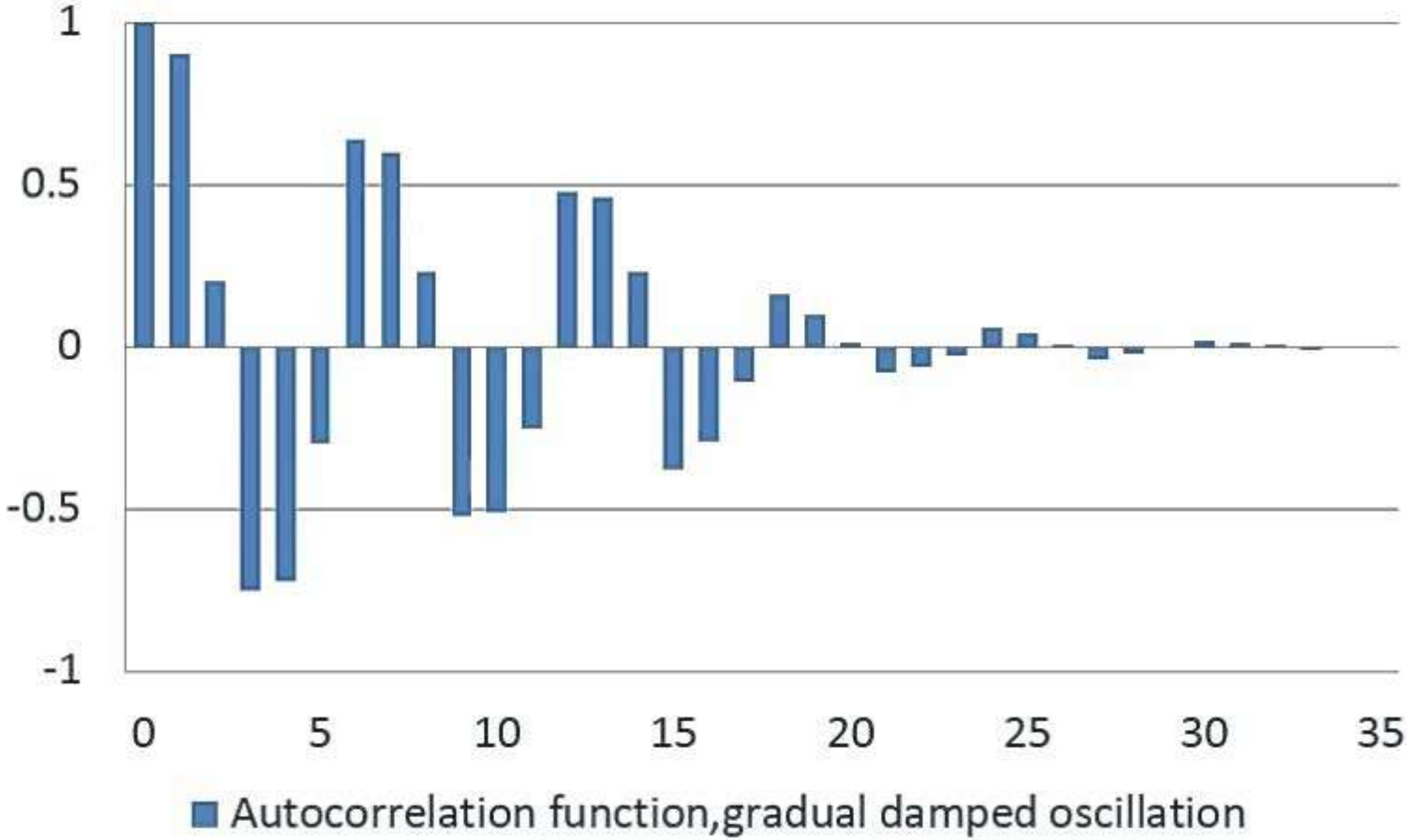
- The four characteristics of autocorrelation function(Graphs)



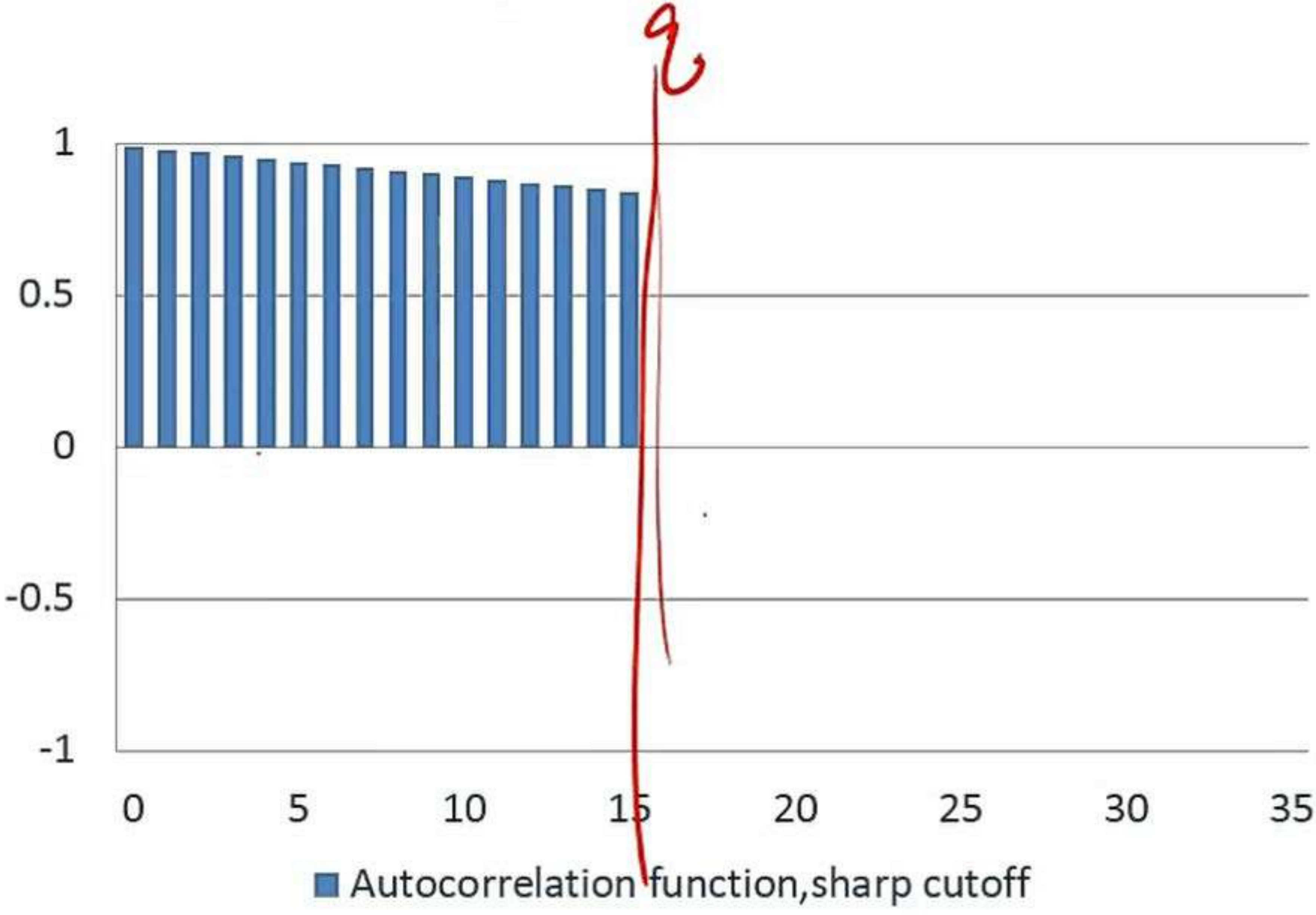
Characterizing Cycles



Characterizing Cycles



Characterizing Cycles



Modeling Cycle

➢ **MA(1) Model** : Moving Average

$$\boxed{y_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1 + \theta L)\varepsilon_t;\ \varepsilon_t \sim WN(0, \sigma^2)}$$

- The current value of the observed series is expressed as a function of current and lagged unobservable shocks.
- It is the very special case of Wold's representation.

➢ **MA(q) Model**

$$\boxed{\begin{gathered} y_t = \varepsilon_t + \theta\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} = \Theta(L)\varepsilon_t;\ \varepsilon_t \sim WN(0, \sigma^2) \\ \Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q \end{gathered}}$$

Modeling Cycle

➤ Characteristic of MA(1)

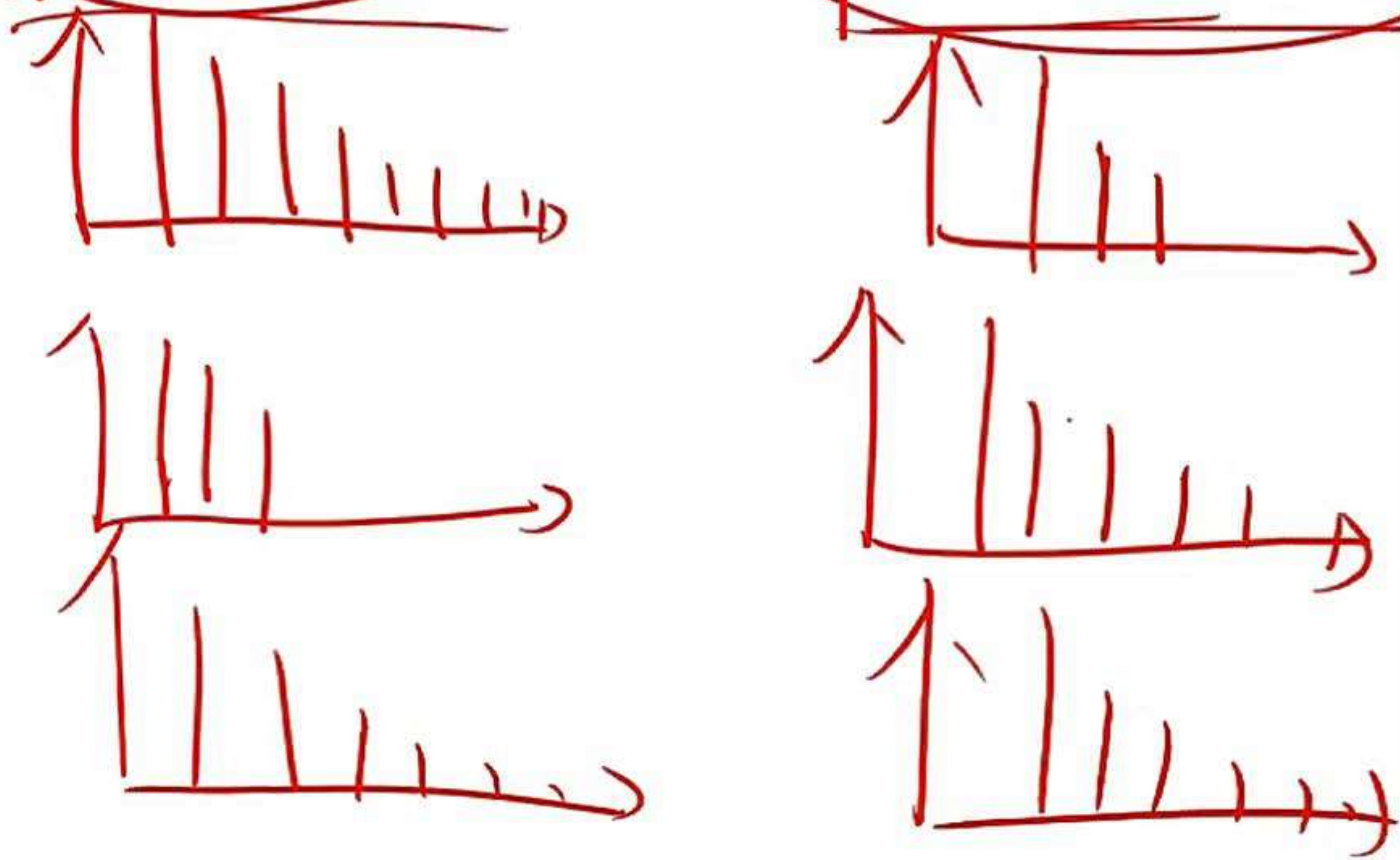
- If the coefficient $|\theta| < 1$, then MA(1) process is **a convergent process.**
- The structure of the MA(1) process, in which only the first lag of the shock appears on the right, forces it to **have a very short memory**, and hence weak dynamics, regardless of the parameter value.
- MA(1) process with parameter θ = 0.95 almost share the same persistence with the process with a parameter of θ = 0.4.
- Autocorrelation graph appears to **have sharp cutoff**.
- Partial Autocorrelation graph appears to have **gradual damped oscillation or one-sided gradual damped.**

autocorrelation — partial correlation

AR

MA

ARMA.



$$y_t = \theta\, y_{t-1}$$

Simulation Methods

一、Monte Carlo 1、过程

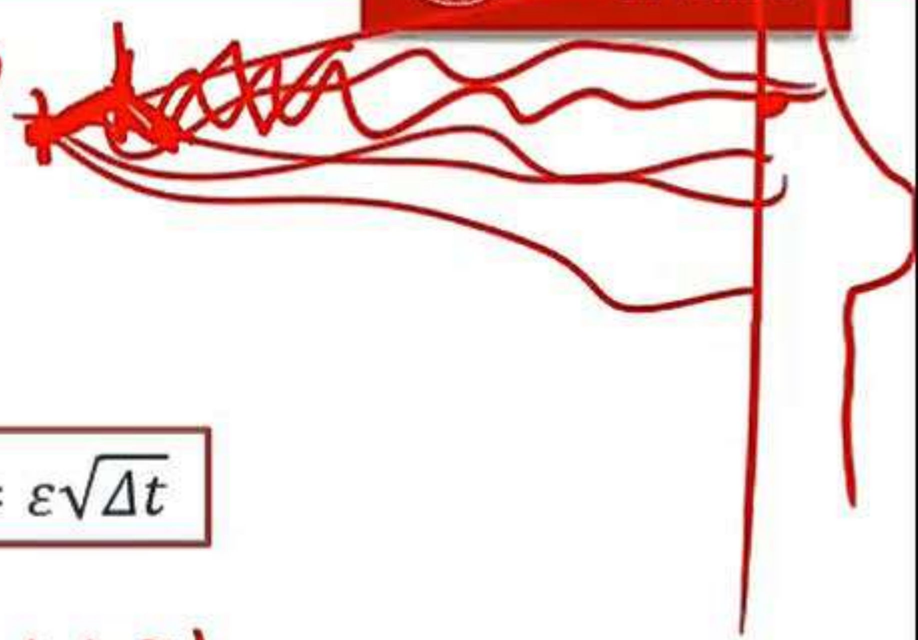
2、公式

3、优缺点

二、Monte Carlo 降低误差的方式

Monte Carlo Simulation

100



➢ **Simulations with one random variable**

- Wiener process 维纳过程

$$\boxed{\Delta z \sim N(0, \Delta t),\ if\ \varepsilon \sim N(0,1),\ then\ \Delta z = \varepsilon\sqrt{\Delta t}}$$

- Generalized wiener process 广义维纳过程

$$\boxed{\Delta x = a\Delta t + b\Delta z}$$

- Ito process 伊藤过程

$$\boxed{\Delta x = a(x,t)\Delta t + b(x,t)\Delta z}$$

- Geometric Brownian motion 几何布朗运动

$$\boxed{\Delta S = \mu S\Delta t + \sigma S\Delta z}$$

漂移项(常数)

Monte Carlo Simulation

- **Simulating a price path**
  - Geometric Brownian Motion (GBM) Model

$$\boxed{dS_t = \mu_t S_t dt + \sigma_t S_t dz}$$

$$\boxed{dz \sim N(0, dt),\ if\ \varepsilon \sim N(0,1),\ then\ \mathrm{dz} = \varepsilon\sqrt{dt}}$$

$\varepsilon \cdot \sqrt{\Delta t}$

  - Where:
    - ✓ St = asset price
    - ✓ dS_t = infinitesimally small price changes
    - ✓ μ_t = constant instantaneous drift term
    - ✓ σ_t = constant instantaneous volatility
    - ✓ d_z = normally distributed random variable (mean = 0, variance = d_t)
  - GBM model is widely used for stock prices and currencies.

Monte Carlo Simulation

- The assumption of GBM is the normal distribution for dS/S = dln(S), that means S follows a lognormal distribution.

ΔS_{t+2}

- **Example**
  - The process of simulation:

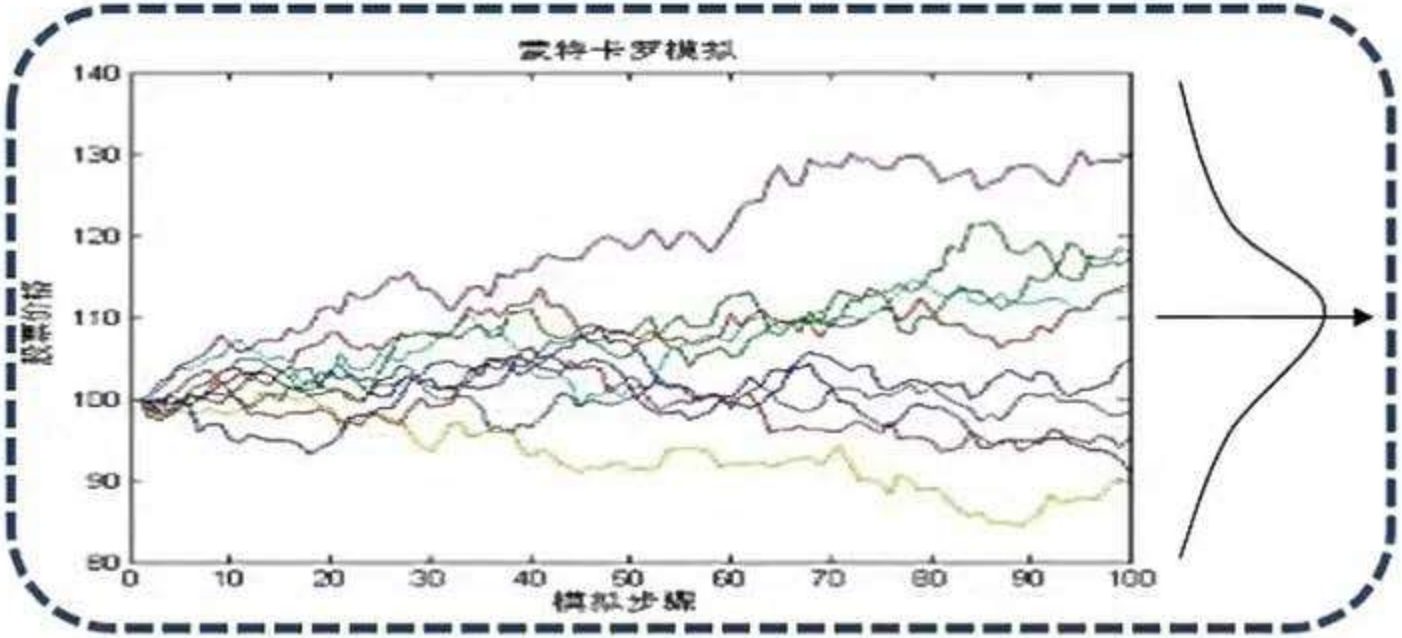
$$\boxed{\Delta S_t = S_t(\mu \Delta t + \sigma \varepsilon \sqrt{\Delta t})}$$

  - Where ε is now a standard normal random variable.
  - Assumption: $\mu = 0$, $\sigma = 0.1$, $S_t = 100$, the result is following (divide the stock price moving into 100 steps, so $\Delta t = 1/100$):

Monte Carlo Simulation

- **Simulating a price path**

Step i	Previous Price S_{t+i-1}	Random Variable ε_i	Increment ΔS	Current Price S_{t+i}
1	100.00	0.199	0.199	100.20
2	100.20	1.665	1.668	101.87
...	...	...	...	...
100	92.47	-1.153	-1.153	91.32



Evaluations of Monte Carlo Simulation

- **Advantages**
  - Flexibility, so they allow financial engineering to price complex financial instruments;
  - Allow risk managers to build the distribution of portfolios that are too complex to model analytically.

Evaluations of Monte Carlo Simulation

GOLDEN FUTURE

- **Disadvantages**

GIGO : garbage in garbage out

- Simulation is good tool for modeling uncertainty, but the outcome is only as good as the inputs we provide to our models: the shape of the distribution, the parameters, and the pricing functions matters.
- As time interval shrinks, the volatility shrinks as well. This implies that large discontinuities cannot occur over short intervals. But in reality, some assets experience discrete jumps.
- Price increments are assumed to have a **normal distribution**. But in reality, price changes have fatter tails and variance of returns can change.
  - ✓ How to select other distributions?

Bootstrapping Method

可以重复抽样

X_1, X_2, X_1+X_2, X_1X_2

- Bootstrapping is related to simulation, but with one crucial difference. With simulation, the data are constructed completely artificially. Bootstrapping, on the other hand, is used to obtain a description of the properties of empirical estimators by using the sample data points themselves, and it involves sampling repeatedly with replacement from the actual data.
- The **advantage** of bootstrapping over the use of analytical results is that it allows the researcher to make inferences without making strong distributional assumptions, since the distribution employed will be that of the actual data.
- **Situations where the bootstrap will be ineffective:**
  - Outliers in the Data
  - Non-Independent Data: Use of the bootstrap implicitly assumes that the data are independent of one another.

Example 1

- Which of the following statements is true regarding the bootstrap simulation method used in VaR estimation?

  I. Bootstrapping uses actual market data.

  II. The bootstrapping method always uses a time horizon based on the time scale of the historical data.

  III. Bootstrapping is based on synthesis of normally distributed random numbers.

A. I only

B. II only

C. I and III

D. I, II and III

Example 1

- Correct Answer: A
  - Bootstrapping uses actual market data. Bootstrapping can be done with data that uses the same time line as the one of interest or shorter term data. Monte Carlo is based on a synthesis of normally distributed random numbers.

Variance Reduction Techniques

二、减少 Monte Carlo 误差方式

① Sampling Variation

- The sampling variation in a Monte Carlo study is measured by the standard error estimate, denoted S_x:

$$S_x = \sqrt{\frac{\text{var}(x)}{N}}$$

1000 ↓ 100倍 100000

$\sigma = 0.1$ ↓ $\frac{1}{\sqrt{100}}$ $\sigma = 0.01$

  ✓ where var(x) is the variance of the estimates of the quantity of interest over the N replications.

- It can be seen from this equation that to reduce the Monte Carlo standard error by a factor of 10, the number of replications must be increased by a factor of 100.
- Consequently, in order to achieve acceptable accuracy, the number of replications may have to be set at an unfeasibly high level. An alternative way to reduce Monte Carlo sampling error is to use a variance reduction technique.

Variance Reduction Techniques

- **Antithetic Variates** 反向变量.
  - The antithetic variate technique involves taking the complement of a set of random numbers and running a parallel simulation on those. For example, if the driving stochastic force is a set of TN(0, 1) draws, denoted u_t, for each replication, an additional replication with errors given by $-u_t$ is also used.
  - It can be shown that the Monte Carlo standard error is reduced when antithetic variates are used.

$\Delta S: \mu \Delta t S \cdots \varepsilon$

$\varepsilon_1 \quad \varepsilon_2 \quad \varepsilon_3 \quad \varepsilon_4 \quad \varepsilon_5$

$-\varepsilon_1 \quad -\varepsilon_2 \quad -\varepsilon_3 \quad -\varepsilon_4 \quad -\varepsilon_5$

Variance Reduction Techniques

➢ Control Variates 控制变量法

- The control variates help to reduce the Monte Carlo variation owing to particular sets of random draws by using the same draws on a related problem whose solution is known. The application of control variates involves employing a variable similar to that used in the simulation, but whose properties are known prior to the simulation.
- Denote the variable whose properties are known by y, and that whose properties are under simulation by x. The simulation is conducted on x and also on y, with the same sets of random number draws being employed in both cases. Denoting the simulation estimates of x and y be $\hat{x}$ and $\hat{y}$, respectively, a new estimate of x can be derived from:

$$x^* = y + (\hat{x} - \hat{y})$$

S_1 100

男朋友 , difference. 蒙特卡罗

10倍 difference S_2' 12

S_1 眼角了

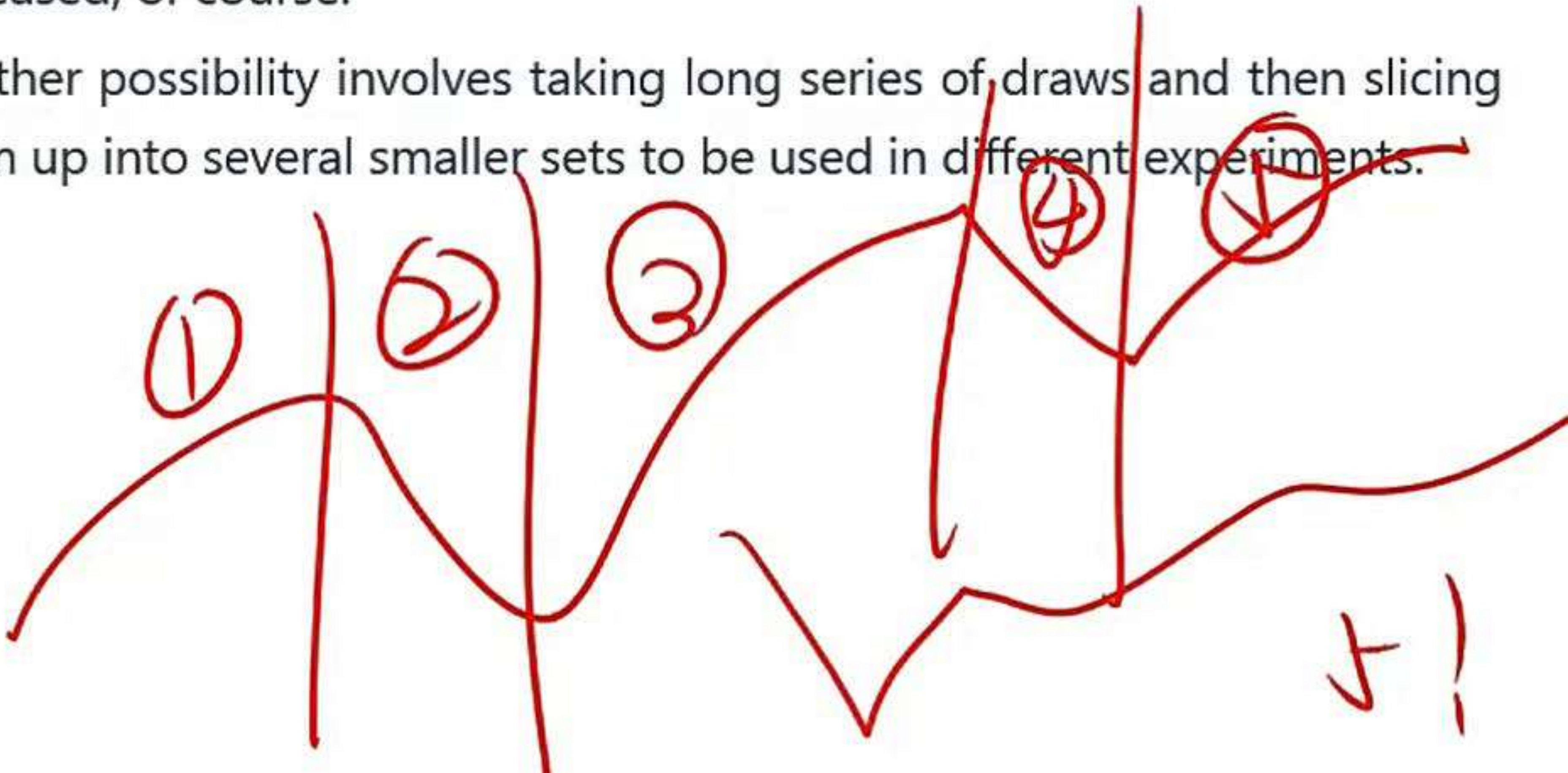
牛

S_1' 10 ,

Variance Reduction Techniques

4 **Random Number Re-Usage across Experiments** 反复利用

- Using the same sets of draws across experiments can greatly reduce the variability of the difference in the estimates across those experiments. However, the accuracy of the actual estimates in each case will not be increased, of course.
- Another possibility involves taking long series of draws and then slicing them up into several smaller sets to be used in different experiments.



Random Number Generation

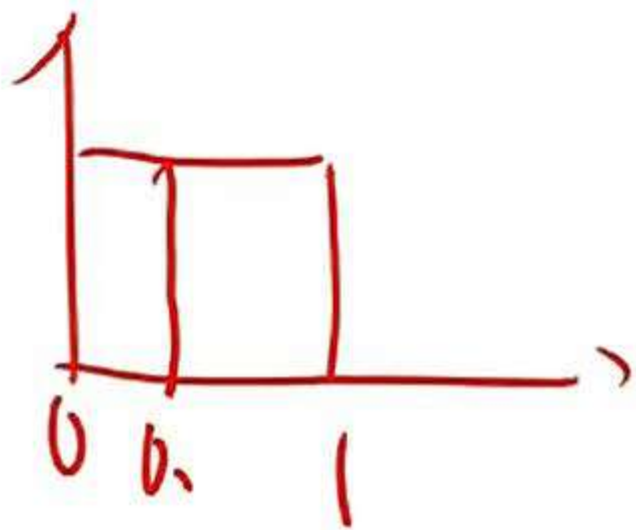
- **Pseudorandom number generators** 伪随机
  - Good uniform random numbers on [0,1] is critical for simulation. Truly random number generation is difficult and time consuming.
  - Use random number generation algorithms that produce streams of numbers that appear to be random. In fact, they are a result of a clearly defined series of calculation steps.
  - It starts with a number called seed and $x_n = g(x_{n-1})$, so if the same seed is used in several simulation, each simulation sequence will contain exactly the same numbers, which is helpful for running fair comparison between different strategies evaluated under uncertainty.

4000万次

Random Number Generation

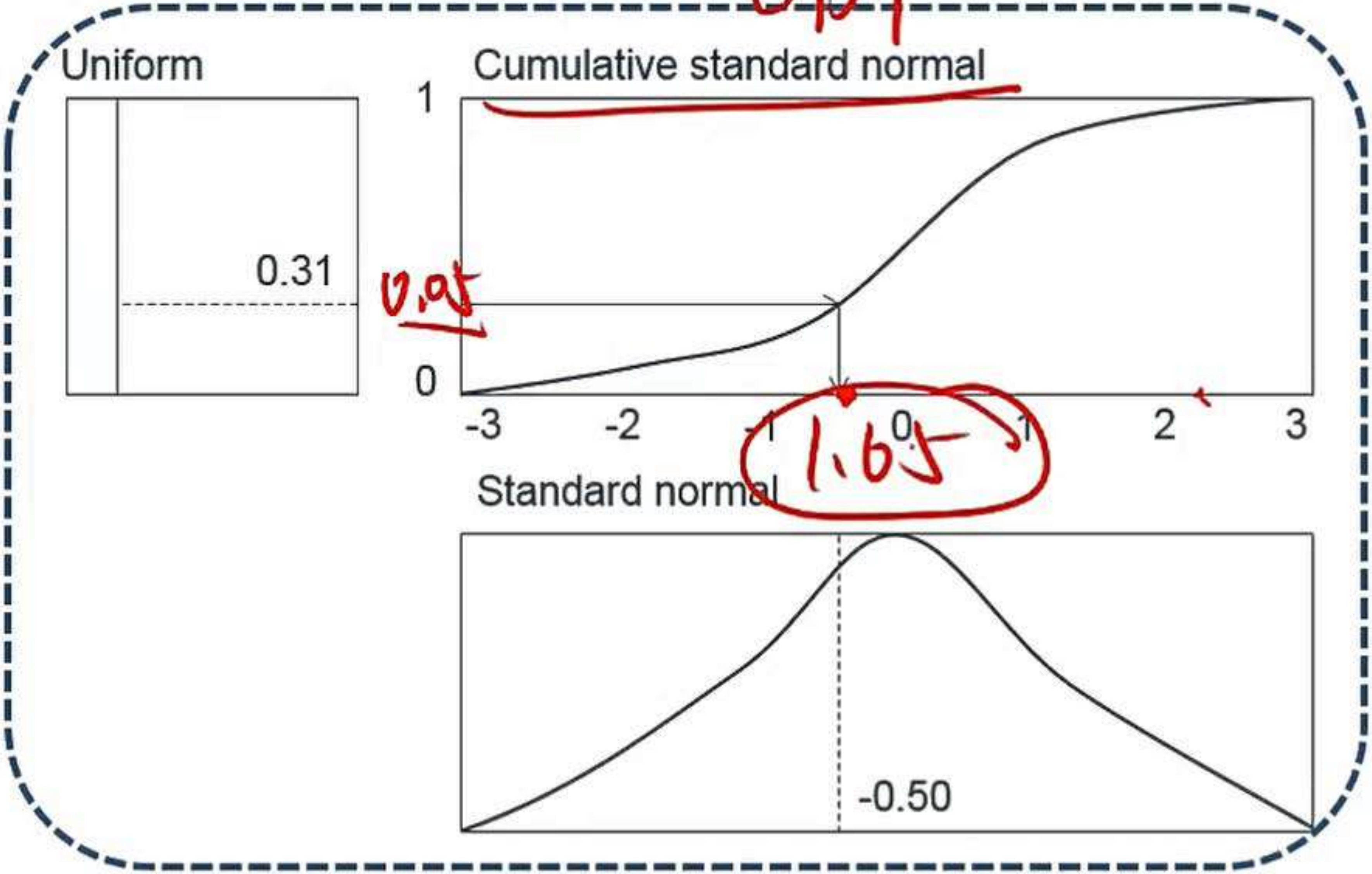
➢ **Standards for effective pseudorandom number generator**

- The numbers in the generated sequence are <u>uniformly distributed</u> between 0 and 1. This can be tested by Chi-Square or K-S test.
- The sequence has a <u>long cycle</u> (takes many iterations before repeating).
- Numbers are not <u>autocorrelated</u>, can be tested by D-W test.



Random Number Generation

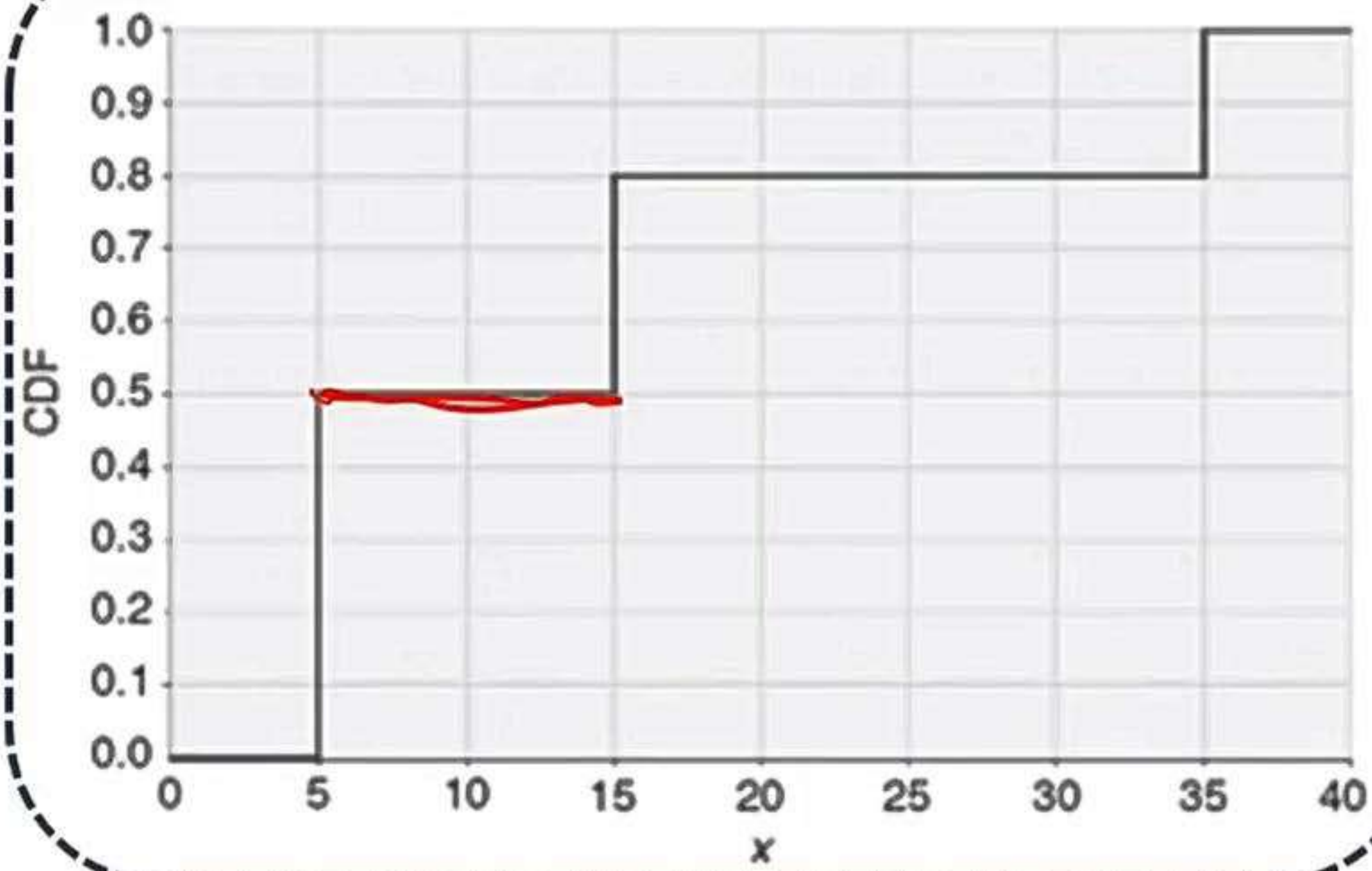
- **Inverse transform method**
  - X ~ U[0,1]



Random Number Generation

- **If it is a discrete distribution:**
  - 5: 50%
  - 15: 30%
  - 35: 20%

CDF



一、ARCH（过程）

二、EWMA

三、GARCH

Estimating Volatilities and Correlations

Estimating Volatilities

ARCH

- **Estimating Volatility**
  - Define σ_n as the volatility of a market variable on day n, as estimated at the end of day n − 1. σ_n^2 as the variance rate.
  - Define S_i as the value of the market variable at the end of day i.
  - Define u_i as the continuously compounded return during day i (between the end of day i − 1 and the end of day i).

$$u_i = \ln \frac{S_i}{S_{i-1}} \qquad \bar{u} = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}$$

$$\sigma_n^2 = \frac{1}{m-1}\sum_{i=1}^{m}(u_{n-i} - \bar{u})^2$$

Estimating Volatilities

ARCH

Estimating Volatility

- Define σ_n as the volatility of a market variable on day n, as estimated at the end of day n − 1. σ_n^2 as the variance rate.
- Define S_i as the value of the market variable at the end of day i.
- Define u_i as the continuously compounded return during day i (between the end of day i − 1 and the end of day i).

$$u_i = \ln \frac{S_i}{S_{i-1}} \qquad \bar{u} = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}$$

$$\sigma_n^2 = \frac{1}{m-1}\sum_{i=1}^{m}(u_{n-i} - \bar{u})^2$$

$X \sim \log \rightarrow \ln(X) \sim$ Normal ✓

Estimating Volatilities

- For the purpose of monitoring daily volatility, we give the following changes:
  - Define u_i as the percentage change in the market variable between the end of day i – 1 and the end of day i.

$$u_i = \frac{S_i - S_{i-1}}{S_{i-1}}$$

  - $\bar{u}$ is assumed to be zero.
  - m – 1 is replaced by m.
- Then we can get a simple formula for the variance rate.

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} u_{n-1}^2$$

Estimating Volatilities

- **Weighting Schemes**
  - The above formula gives equal weight to u_{n-1}^2, u_{n-2}^2, ……, u_{n-m}^2
  - Our objective is to estimate the current level of volatility, so we give more weight to recent data.

$$\sigma_n^2 = \sum_{i=1}^{m} \alpha_i u_{n-1}^2 \qquad \alpha_i < \alpha_j \text{ where i > j} \qquad \sum_{i=1}^{m} \alpha_i = 1$$

  - If the objective is to generate a greater influence on recent observations, then the α's will decline in value for older observations.

Estimating Volatilities

➢ ARCH Model

- Adding a long-run average variance rate and be given a weight.

$$\boxed{\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u^2{}_{n-1}} \quad \boxed{\gamma + \sum_{i=1}^{m} \alpha_i = 1}$$

✓ Where V_L is the long-run variance rate and γ is the weight assigned to V_L.

- Defining $\omega = \gamma V_L$, then the model can be written:

$$\boxed{\sigma_n^2 = \omega + \sum_{i=1}^{m} \alpha_i u^2{}_{n-1}}$$

Estimating Volatilities

➢ EWMA Model

- In an exponentially weighted moving average model, as time goes by, the weight assigned to α_i is exponentially decreasing.

$$\boxed{\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2}$$

- The estimate, σ_n , of the volatility for day n (made at the end of day n-1) is calculated from σ_{n-1} (the estimate made at the end of day n-2 of the volatility for day n-1) and u_{n-1} (the most recent daily percentage change).
- High values of λ will minimize the effect of daily percentage returns, whereas low values of λ will tend to increase the effect of daily percentage returns on the current volatility estimate.

Estimating Volatilities

➢ **EWMA Model** ①预测 σ

- In an exponentially weighted moving average model, as time goes by, the weight assigned to α_i is exponentially decreasing.

②公式 $$\boxed{\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2}$$

- The estimate, σ_n, of the volatility for day n (made at the end of day n-1) is calculated from σ_{n-1} (the estimate made at the end of day n-2 of the volatility for day n-1) and u_{n-1} (the most recent daily percentage change).
- High values of λ will minimize the effect of daily percentage returns, whereas low values of λ will tend to increase the effect of daily percentage returns on the current volatility estimate.

Estimating Volatilities

➢ **EWMA Model** ①预测 σ

0.94

- In an exponentially weighted moving average model, as time goes by, the weight assigned to α_i is exponentially decreasing.

②公式 $$\boxed{\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2}$$

- The estimate, σ_n , of the volatility for day n (made at the end of day n-1) is calculated from σ_{n-1} (the estimate made at the end of day n-2 of the volatility for day n-1) and u_{n-1} (the most recent daily percentage change).
- High values of λ will minimize the effect of daily percentage returns, whereas low values of λ will tend to increase the effect of daily percentage returns on the current volatility estimate.

Estimating Volatilities

➢ **EWMA Model** ①预测 σ

0.94

- In an exponentially weighted moving ~~average model~~, as time goes by, the weight assigned to α_i is exponentially decreasing.

③ u_{n-1}^2 的 weight ②公式

$$\boxed{\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2}$$

- The estimate, σ_n, of the volatility for day n (made at the end of day n-1) is calculated from σ_{n-1} (the estimate made at the end of day n-2 of the volatility for day n-1) and u_{n-1} (the most recent daily percentage change).
- High values of λ will minimize the effect of daily percentage returns, whereas low values of λ will tend to increase the effect of daily percentage returns on the current volatility estimate.

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1-\lambda)\mu_{n-1}^2$$

$$\sigma_{n-1}^2 = \lambda \sigma_{n-2}^2 + (1-\lambda)\mu_{n-2}^2$$

$$\sigma_n^2 = \lambda\left[\lambda\cdot\sigma_{n-2}^2 + (1-\lambda)\mu_{n-2}^2\right] + (1-\lambda)\mu_{n-1}^2$$

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1-\lambda)\mu_{n-1}^2$$

$$\sigma_{n-1}^2 = \lambda \sigma_{n-2}^2 + (1-\lambda)\mu_{n-2}^2$$

$$\sigma_n^2 = \lambda\left[\lambda \cdot \sigma_{n-2}^2 + \lambda(1-\lambda)\mu_{n-2}^2\right] + (1-\lambda)\mu_{n-1}^2$$

Estimating Volatilities

➢ **Using EWMA To Forecast Future Volatility**

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$$

$$\sigma_n^2 = \lambda[\lambda\sigma_{n-2}^2 + (1-\lambda)u_{n-2}^2] + (1-\lambda)u_{n-1}^2 = (1-\lambda)(u_{n-1}^2 + \lambda u_{n-2}^2) + \lambda^2\sigma_{n-2}^2$$

$$\sigma_n^2 = (1-\lambda)(u_{n-1}^2 + \lambda u_{n-2}^2) + \lambda^2\sigma_{n-2}^2 = (1-\lambda)(u_{n-1}^2 + \lambda u_{n-2}^2 + \lambda^2 u_{n-3}^2) + \lambda^3\sigma_{n-3}^2$$

$\lambda^2(1-\lambda)$

……

$$\sigma_n^2 = (1-\lambda)\sum_{i=1}^{m}\lambda^{i-1}u_{n-i}^2 + \lambda^m\sigma_{n-m}^2$$

➢ **从现在算起，i天前收益率平方的权重记为α_i，那么$\alpha_i = (1-\lambda)\lambda^{i-1}$.**

u_{n-i}^2 $\lambda^{i-1}(1-\lambda)$

$\lambda^0(1-\lambda)$

Estimating Volatilities

- **EWMA Model** ①预测 σ

0.94

  - In an exponentially weighted moving average model, as time goes by, the weight assigned to α_i is exponentially decreasing.

②公式

$$\boxed{\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2}$$

③ u_{n-i}^2 的 weight

④ EWMA 是 a special case of GARCH

  - The estimate, σ_n , of the volatility for day n (made at the end of day n-1) is calculated from σ_{n-1} (the estimate made at the end of day n-2 of the volatility for day n-1) and u_{n-1} (the most recent daily percentage change).
  - High values of λ will minimize the effect of daily percentage returns, whereas low values of λ will tend to increase the effect of daily percentage returns on the current volatility estimate.

⑤ EWMA：未考虑长期方差项

Estimating Volatilities

- Example
  - Suppose that λ is 0.90, the volatility estimated for a market variable for day n – 1 is 1% per day, and during day n – 1 the market variable increased by 2%.

$\sigma_{n-1} = 1\%$

$u_{n-1} \quad 2\%$

$$\sigma_{n-1}^2 = 0.01^2 = 0.0001$$

$$\mu_{n-1}^2 = 0.02^2 = 0.0004$$

$$\sigma_n^2 = 0.9 \times 0.0001 + 0.1 \times 0.0004 = 0.00013$$

Estimating Volatilities

- Given λ of 0.94, under an infinite series, what is the weight assigned to the seventh prior daily squared return?
  - A. 4.68%
  - B. 4.40%
  - C. 4.14%
  - D. 3.89%

$$\boxed{\quad}\, u_{n-7}^2$$

$$\lambda^{i-1}(1-\lambda)$$

$$\lambda^{6}(1-\lambda)$$

- Correct Answer: C

$$\text{Weight} = 0.94^6 \times (1 - 0.94) = 4.14\%$$

Estimating Volatilities

- **The Attractive Feature Of EWMA Approach**
  - Relatively little data needs to be stored.
  - We need only remember the current estimate of the variance rate and the most recent observation on the value of the market variable.
  - Tracks volatility changes. The value of λ governs how responsive the estimate of the daily volatility is to the most recent daily percentage change.
  - Risk Metrics uses $\lambda = 0.94$ for daily volatility forecasting.

Estimating Volatilities

三 ➢ **GARCH(1, 1) Model** ①预测 σ

- In GARCH(1, 1), σ_n^2 is calculated from a long-run average variance rate VL, as well as from σn-1 and un-1. The equation for GARCH(1, 1) is:

③ $\alpha + \beta + \gamma = 1$

②公式 $$\boxed{\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2}$$ ④ $\alpha + \beta < 1$

- EWMA model is a particular case of GARCH(1, 1), where $\gamma = 0$, $\alpha = 1 - \lambda$, $\beta = \lambda$.

- The "(1, 1)" in GARCH(1, 1) indicates that σ_n^2 is based on the most recent observation of u_{n-1} and the most recent estimate of the variance rate.

- Setting $\omega = \gamma V_L$, the GARCH(1, 1) model can also be written:

$$\boxed{\sigma_n^2 = \omega + \alpha u_{n-1} + \beta \sigma_{n-1}}$$

Estimating Volatilities

三 GARCH(1, 1) Model ①预测 σ

- In GARCH(1, 1), $\sigma_n^{\ 2}$ is calculated from a long-run average variance rate VL, as well as from σn-1 and un-1. The equation for GARCH(1, 1) is:

③ $\alpha+\beta+\gamma=1$

②公式 $$\boxed{\sigma_n^{\ 2} = \gamma V_L + \alpha u_{n-1}^{\ 2} + \beta\sigma_{n-1}^{\ 2}}$$ ④ $\alpha+\beta<1$

- EWMA model is a particular case of GARCH(1, 1), where $\gamma = 0$, $\alpha = 1 - \lambda$, $\beta = \lambda$.

⑤ $\alpha+\beta$ persistence 持续

- The "(1, 1)" in GARCH(1, 1) indicates that σ_n^2 is based on the most recent observation of u_{n-1} and the most recent estimate of the variance rate.

- Setting $\omega = \gamma V_L$, the GARCH(1, 1) model can also be written:

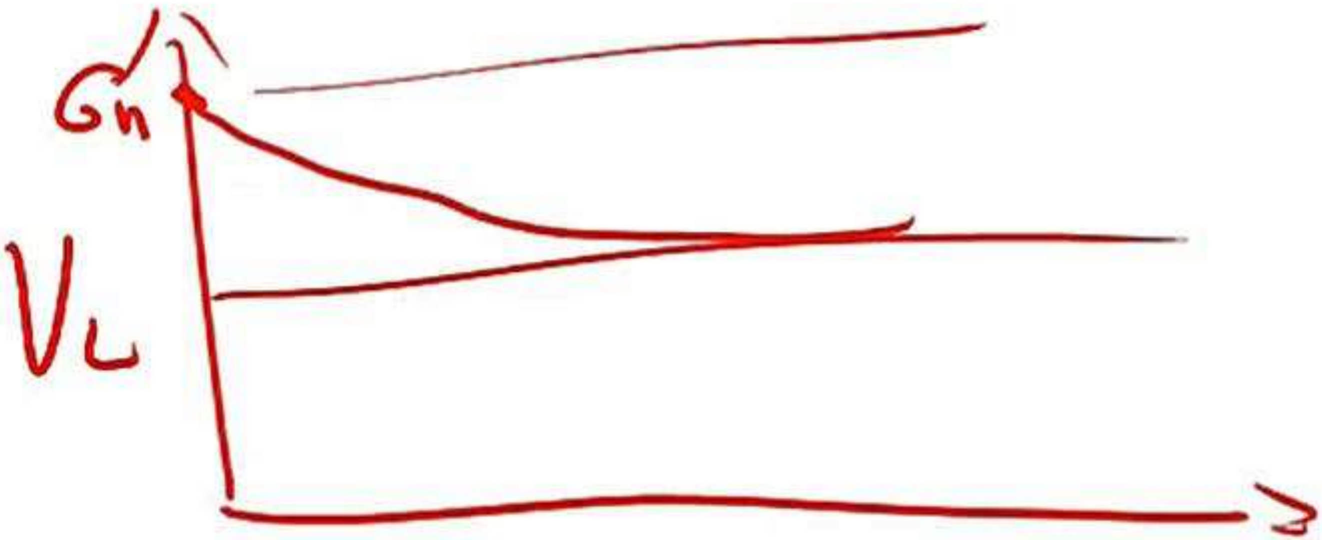
$$\boxed{\sigma_n^{\ 2} = \omega + \alpha u_{n-1} + \beta\sigma_{n-1}}$$

GARCH σ_n

① $\cdot$ $\alpha+\beta$ persistence

$\underline{\alpha+\beta} \rightarrow 1$ time longer

② mean reverting

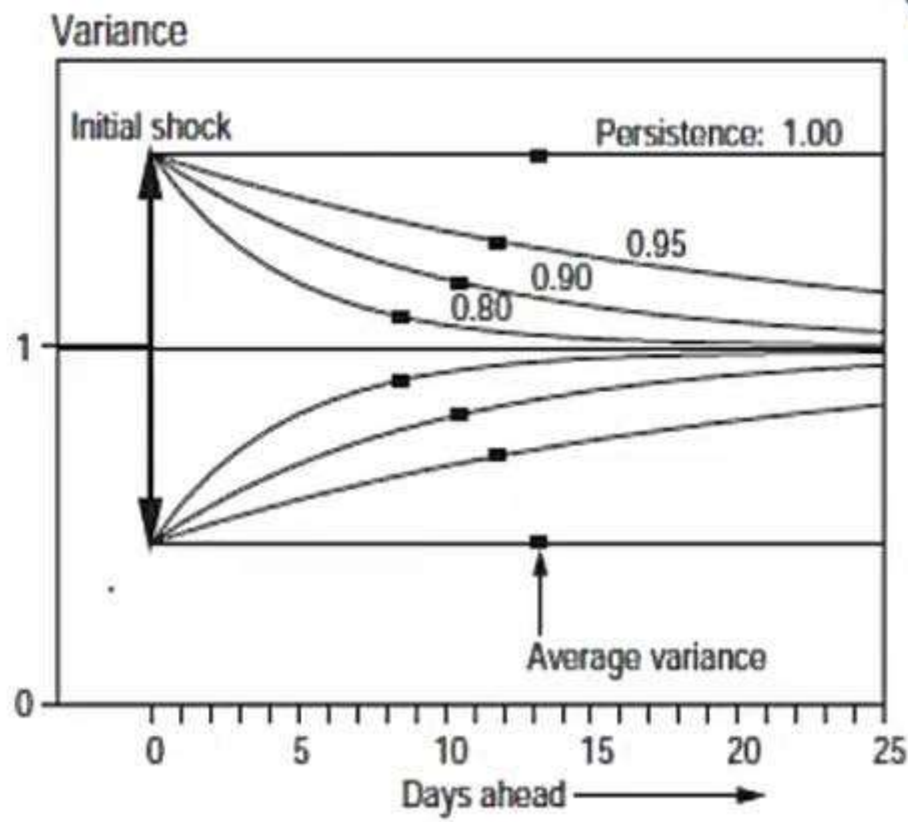


Estimating Volatilities

➢ **GARCH(1, 1) Model**

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

- Persistence: $\alpha + \beta$
- It defines the speed at which shocks to the variance revert to their long-run values.
- The higher the persistence (given that it is less than one), the longer it will take to revert to the mean following a shock or large movement.



Shocks to a GARCH Process

Estimating Volatilities

$$V_L = \frac{0.000002}{0.01}$$

➢ **Example**

- Suppose that a GARCH(1, 1) model is estimated from daily data as:

$$\sigma_n^{\,2} = 0.000002 + 0.13u_{n-1}^{\,2} + 0.86\,\sigma_{n-1}^{\,2}$$

- This corresponds to $\alpha = 0.13$, $\beta = 0.86$, $\omega = 0.000002$
- Because $\gamma = 1 - \alpha - \beta$, it follows that $\gamma = 0.01$
- Because $\omega = \gamma V_L$, it follows that $V_L = 0.0002$. This corresponds to a volatility of $0.014 = 1.4\%$ per day.
- Suppose that the estimate of volatility on day $n - 1$ is 1.6% per day, and that on day $n - 1$ the market variable decreased by 1%. Then:

$$\sigma_n^{\,2} = 0.000002 + 0.13 \times 0.01^2 + 0.86 \times 0.016^2 = 0.00023516$$

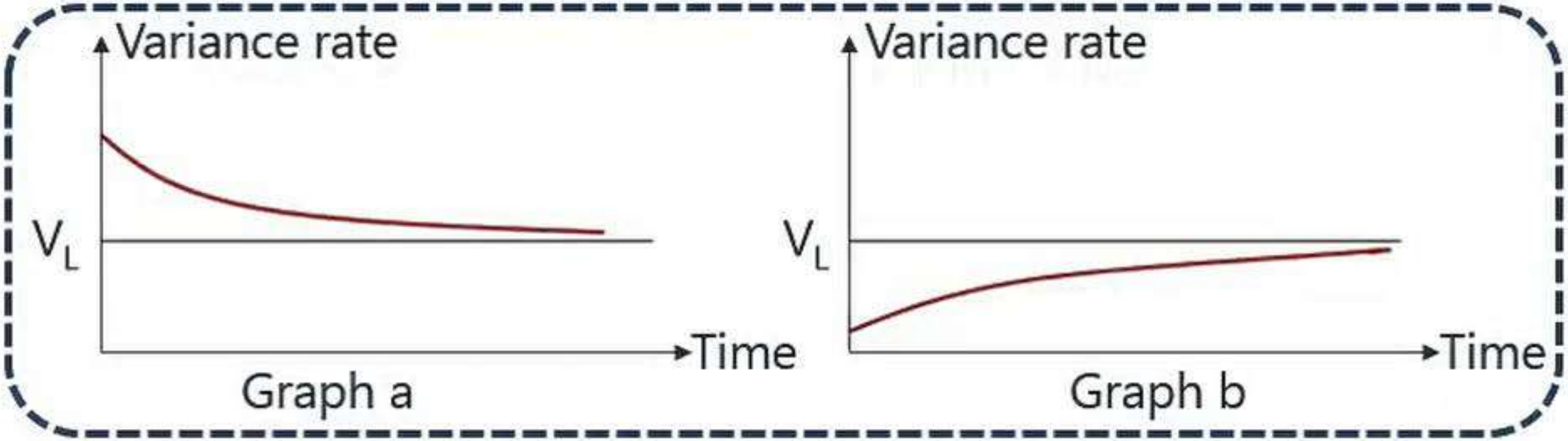
- The new estimate of the volatility is therefore $0.0153 = 1.53\%$ per day.

Estimating Volatilities

Choosing between the models

- GARCH(1, 1) is theoretically more appealing than EWMA model.
- In practice, variance rates tend to be mean reverting. The GARCH(1, 1) model incorporates mean reversion, whereas the EWMA model does not.

Mean Reversion



- (a) current variance rate is above long-term variance rate.
- (b) current variance rate is below long-term variance rate.

Example 1

$V_L = \frac{w}{1-\alpha-\beta} = \frac{0.000005}{0.03}$

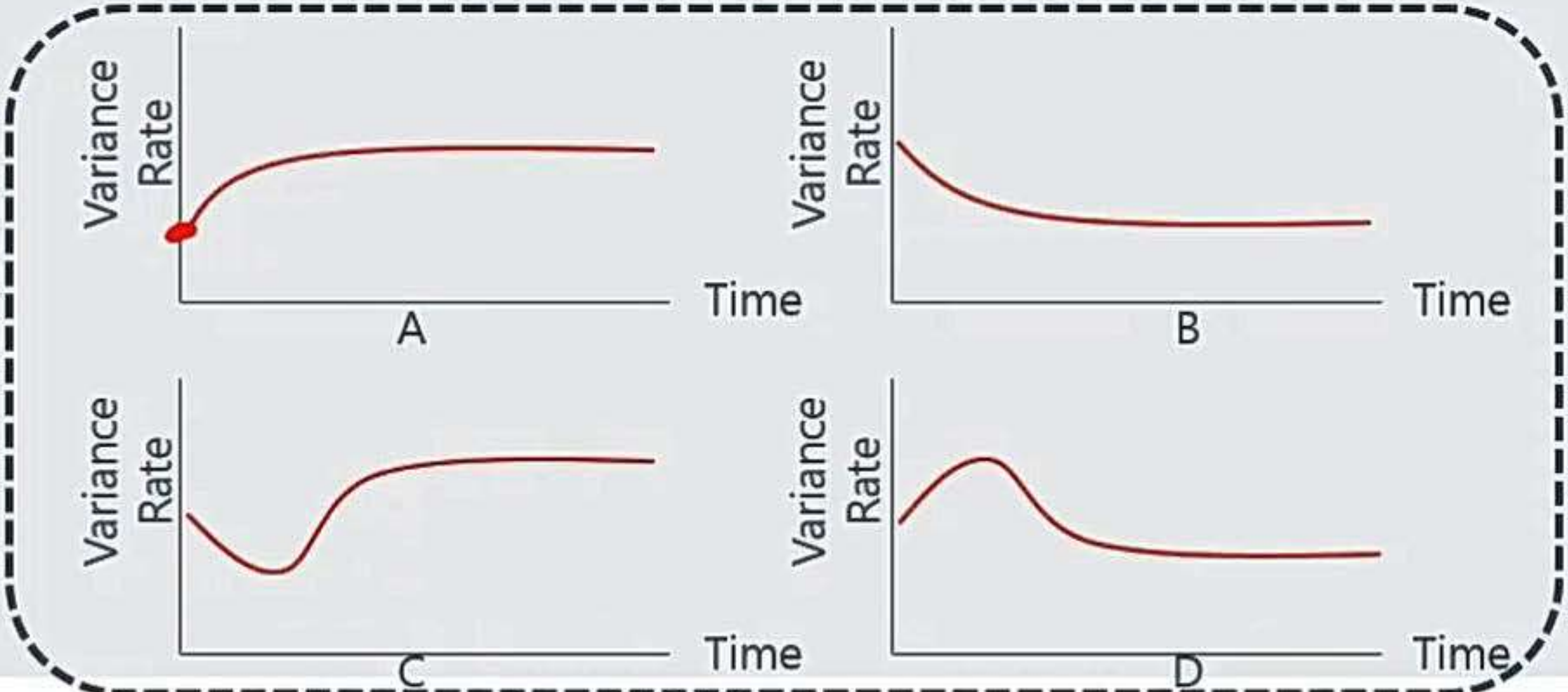
- The following GARCH(1,1) model is used to forecast the daily return variance of an asset:

$$\sigma_n^2 = 0.000005 + 0.05u_{n-1}^2 + 0.92\sigma_{n-1}^2$$

$\sigma_{n-1}^2 = 5\%^2$

$u_{n-1}^2 = 2\%^2$

Suppose the estimate of the volatility today is 5.0% and the asset return is -2.0%. The resulting volatility term structure from this GARCH model is most likely to be represented:



- Using a daily Risk Metrics EWMA model with a decay factor λ = 0.95 to develop a forecast of the conditional variance, which weight will be applied to the return that is four days old?

  A. 0.000

  B. 0.043

  C. 0.048

  D. 0.950

$\lambda^{i-1}(1-\lambda)$

$\lambda^{3}(1-\lambda)$

- Correct Answer : B

Example 3

- An ~~investment bank~~ uses the Exponentially Weighted Moving Average (EWMA) technique with λ of 0.9 to model the daily volatility of a security. The current estimate of the daily volatility is 1.5%. The closing price of the security is USD 20 yesterday and USD 18 today. Using continuously compounded returns, what is the updated estimate of the volatility?

  A. 3.62%

  B. 1.31%

  C. 2.96%

  D. 5.44%

$\sigma_{n-1} = 1.5\%$ $\qquad \ln \frac{18}{20} = -$

$\sigma_n^2 = ?$

- Correct Answer : A

Example 4

B

- Which of the following GARCH models will take the shortest time to revert to its long-run value?

$\alpha+\beta \to 1$ long

A. $\sigma_n^2 = 0.05 + 0.03u_{n-1}^2 + 0.96\sigma_{n-1}^2$ 0.99

B. $\sigma_n^2 = 0.03 + 0.02u_{n-1}^2 + 0.95\sigma_{n-1}^2$ 0.97 $\alpha+\beta$ 最小

C. $\sigma_n^2 = 0.02 + 0.01u_{n-1}^2 + 0.97\sigma_{n-1}^2$ 0.98

D. $\sigma_n^2 = 0.01 + 0.01u_{n-1}^2 + 0.98\sigma_{n-1}^2$ 0.99

- Correct Answer： B

Example 5

- The following GARCH(1,1) model is used to forecast the daily return variance of an asset:

$$\sigma_n^2 = 0.000005 + 0.05u_{n-1}^2 + 0.92\sigma_{n-1}^2$$

Suppose the estimate of the volatility today is 5.0% and the asset return is -2.0%. What is the estimate of the long-run average volatility per day?

A. 1.29%

B. 1.73%

C. 1.85%

D. 1.91%

A

$\sigma_{n-1}^2 = 5\%^2 \quad u_{n-1}^2 = 2\%$

$V_L = \frac{w}{1-\alpha-\beta} = \frac{0.000005}{0.03}$

- Correct Answer : A

总结：

EWMA.

① 用途：forecast volatility

② 公式：$\sigma_n^2 = \lambda \cdot \sigma_{n-1}^2 + (1-\lambda) u_{n-1}^2$.

③ i天前 u_{n-i}^2 前的系数：$\lambda^{i-1}(1-\lambda)$

④ EWMA is a special case of GARCH.

⑤ EWMA 未考虑 V_L.

GARCH

① 用途：forecast volatility

② 公式：$\sigma_n^2 = \omega + \alpha \cdot u_{n-1}^2 + \beta \sigma_{n-1}^2$.

$\sigma_n^2 = \gamma \cdot V_L + \alpha \cdot u_{n-1}^2 + \beta \cdot \sigma_{n-1}^2$.

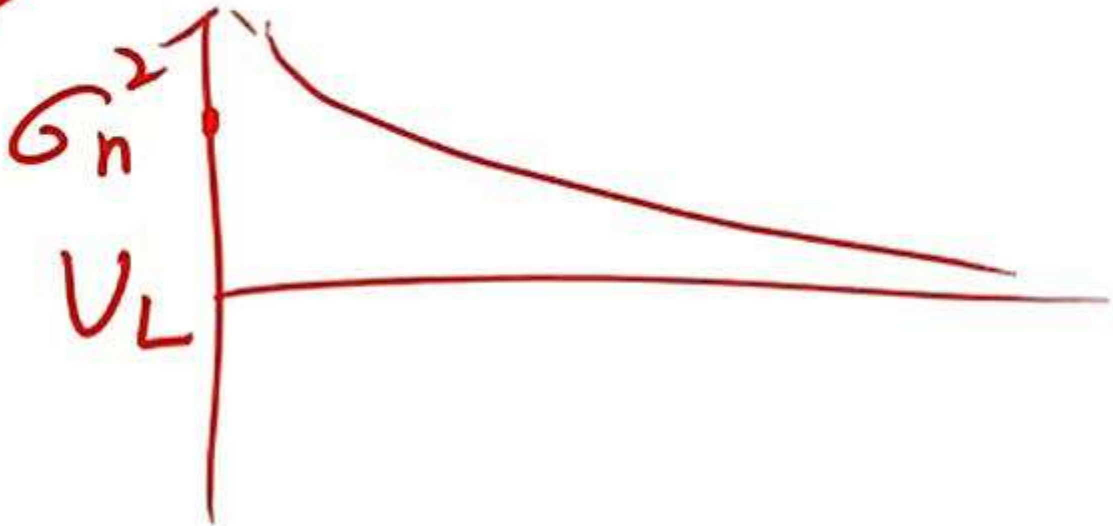
③ $\alpha + \beta + \gamma = 1$

④ $\alpha + \beta < 1 \Rightarrow$ GARCH 收敛

⑤ $\alpha + \beta$：persistence. 持续因子.

$\alpha + \beta \to 1$ mean reverting 了 longer

⑥ GARCH mean reverting.



upward-sloping

downward-sloping

Correlations and Copulas

一、ρ 的求解 ✓

二、矩阵基本分解

三、copula. ✓

四. Tail dependence.

Estimating Correlations

- **Estimating Correlations**

$$\hat{\rho}_{XY,n} = \frac{Cov_n}{\boxed{\sigma_{x,n}}\sigma_{y,n}}$$

- For EWMA model : $\sigma_n^2 = \lambda \cdot \sigma_{n-1}^2 + (1-\lambda) u_{n-1}^2$

$$Cov_n = \lambda Cov_{n-1} + (1-\lambda)x_{n-1}y_{n-1}$$

- For GARCH(1,1) model : $\sigma_n^2 = \omega + \alpha \cdot u_{n-1}^2 + \beta \sigma_{n-1}^2$

$$Cov_n = \omega + \alpha x_{n-1}y_{n-1} + \beta Cov_{n-1}$$

Example 1

- Suppose that $\lambda = 0.95$ and that the estimate of the correlation between two variable X and Y on day $n - 1$ is 0.6. Suppose further that the estimate of the volatilities for the X and Y on day $n - 1$ are 1% and 2%, respectively. From the relationship between correlation and covariance, the estimate of the covariance between the X and Y on day $n - 1$ is:

$$Cov_{n-1} = 0.6 \times 0.01 \times 0.02 = 0.00012$$

- Suppose that the percentage changes in X and Y on day $n - 1$ are 0.5% and 2.5%, respectively. The variance and covariance for day n would be updated as follows:

$$\sigma_{x,n}{}^2 = 0.95 \times 0.01^2 + 0.05 \times 0.005^2 = 0.00009625$$

$$\sigma_{y,n}{}^2 = 0.95\,\sigma^2 \times 0.02^2 + 0.05 \times 0.025^2 = 0.00041125$$

$$Cov_n = 0.95 \times 0.00012 + 0.05 \times 0.005 \times 0.025 = 0.0001202$$

$$\hat{\rho}_{XY} = \frac{Cov_n}{\sigma_{x,n}\sigma_{y,n}} = \frac{0.00012025}{\sqrt{0.0009625} \times \sqrt{0.00041125}} = 0.6044$$

Example 1

$\lambda = 0.95$ $\rho_{n-1} = 0.6$ $\sigma_{n-1,x} = 1\%$

$\sigma_{n-1,y} = 2\%$

- Suppose that λ = 0.95 and that the estimate of the correlation between two variable X and Y on day n – 1 is 0.6. Suppose further that the estimate of the volatilities for the X and Y on day n – 1 are 1% and 2%, respectively. From the relationship between correlation and covariance, the estimate of the covariance between the X and Y on day n – 1 is:

$\rho_n = \frac{cov_n}{\sigma_{x,n}\sigma_{y,n}}$

$$\mathrm{Cov}_{n-1} = \underset{\rho}{0.6} \times \underset{\sigma_x}{0.01} \times \underset{\sigma_y}{0.02} = 0.00012$$

- Suppose that the percentage changes in X and Y on day n – 1 are 0.5% and 2.5%, respectively. The variance and covariance for day n would be updated as follows:

$u_{n-1,x} = 0.5\%$

$$\sigma_{x,n}{}^2 = 0.95 \times 0.01^2 + 0.05 \times 0.005^2 = 0.00009625$$

$$\sigma_{y,n}{}^2 = 0.95\,\sigma^2 \times 0.02^2 + 0.05 \times 0.025^2 = 0.00041125$$

$u_{n-1,y} = 2.5\%$

$$Cov_n = 0.95 \times 0.00012 + 0.05 \times 0.005 \times 0.025 = 0.0001202$$

$$\sigma_{n,x}^2 = \lambda \cdot \sigma_{n-1,x}^2 + (1-\lambda)\, u_{n-1,x}^2$$

$$\hat{\rho}_{XY} = \frac{\mathrm{Cov}_n}{\sigma_{x,n}\sigma_{y,n}} = \frac{0.00012025}{\sqrt{0.00009625} \times \sqrt{0.00041125}} = 0.6044$$

$$cov_n = \lambda \cdot cov_{n-1} + (1-\lambda)\, u_{n-1,x} \cdot u_{n-1,y}$$

Cholesky Factorization

- **Simulate correlated random variables**

$$\Delta S_{1,t} = S_{1,t-1}\mu_1\Delta t + S_{1,t-1}\sigma_1\varepsilon_{1,t}\sqrt{\Delta t}$$

$$\Delta S_{2,t} = S_{2,t-1}\mu_2\Delta t + S_{2,t-1}\sigma_2\varepsilon_{2,t}\sqrt{\Delta t}$$

- To account for correlations between variables, we start with a set of independent variables η of unit variance, which then are transformed into the ε. In a two variable setting, we construct

$$\boxed{\varepsilon_1 = \eta_1}$$

$$\boxed{\varepsilon_2 = \rho\eta_1 + (1-\rho^2)^{1/2}\eta_2}$$

$$\boxed{\begin{gathered} V(\varepsilon_1) = V(\varepsilon_2) = 1 \\ \operatorname{cov}(\varepsilon_1, \varepsilon_2) = \rho \\ V(\varepsilon) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = R \end{gathered}}$$

Copula

➢ **Drawbacks of using correlation to measure dependence.**

- Correlation is a good measure of dependence when random variables are distributed as multivariate elliptical (e.g., normal, student's).
- However, correlation is only defined if variance is finite. There might be a problem for **non-elliptical** distributions. e.g. Levy distribution can have infinite variance.
- Correlation measures the linear relationship of two variables. If risks are independent → zero correlation, however zero correlation does not imply independence.

Copula

➢ **Drawbacks of using correlation to measure dependence.**

- Correlation is a good measure of dependence when random variables are distributed as multivariate elliptical (e.g., normal, student's).

  椭圆分布 $\begin{cases} \sigma \neq 0 \\ \sigma \text{可以被计算} \end{cases}$

- However, correlation is only defined if variance is finite. There might be a problem for **non-elliptical** distributions. e.g. Levy distribution can have infinite variance.
- Correlation measures the linear relationship of two variables. If risks are independent → zero correlation, however zero correlation does not imply independence.

Copula

- **Introduction of Copula**
  - A "copula" is Latin noun which means 'a link, tie or bond'
- **More about Copula:**
  - When the two variables are independent, the joint density is simply the product of the marginal densities.
  - It is rarely the case, however, that financial variables are independent. Dependencies can be modeled by a **function called the copula**, which links, or attaches, marginal distributions into a joint distribution.

Copula

- **Copula function**
  - Formally, the copula is a function of the marginal distributions F (x), plus some parameters, θ, that are specific to this function (and not to the marginals). In the bivariate case, it has two arguments

$$c_{12}[F_1(x_1), F_2(x_2);\ \theta]$$

    - ✓ The link between the joint and marginal distribution is made explicit by Sklar's theorem, which states that, for any joint density, there exists a copula that links the marginal densities

$$f_{12}(x_1, x_2) = f_1(x_1) \times f_2(x_2) \times c_{12}[F_1(x_1), F_2(x_2);\ \theta]$$

      - ◆ With independence, the copula function is a constant always equal to one.

Copula

➢ **Various Types of Copulas**

- Gaussian Copula: most common correlation structure where U_1 and U_2 are assumed to follow bivariate Normal distribution

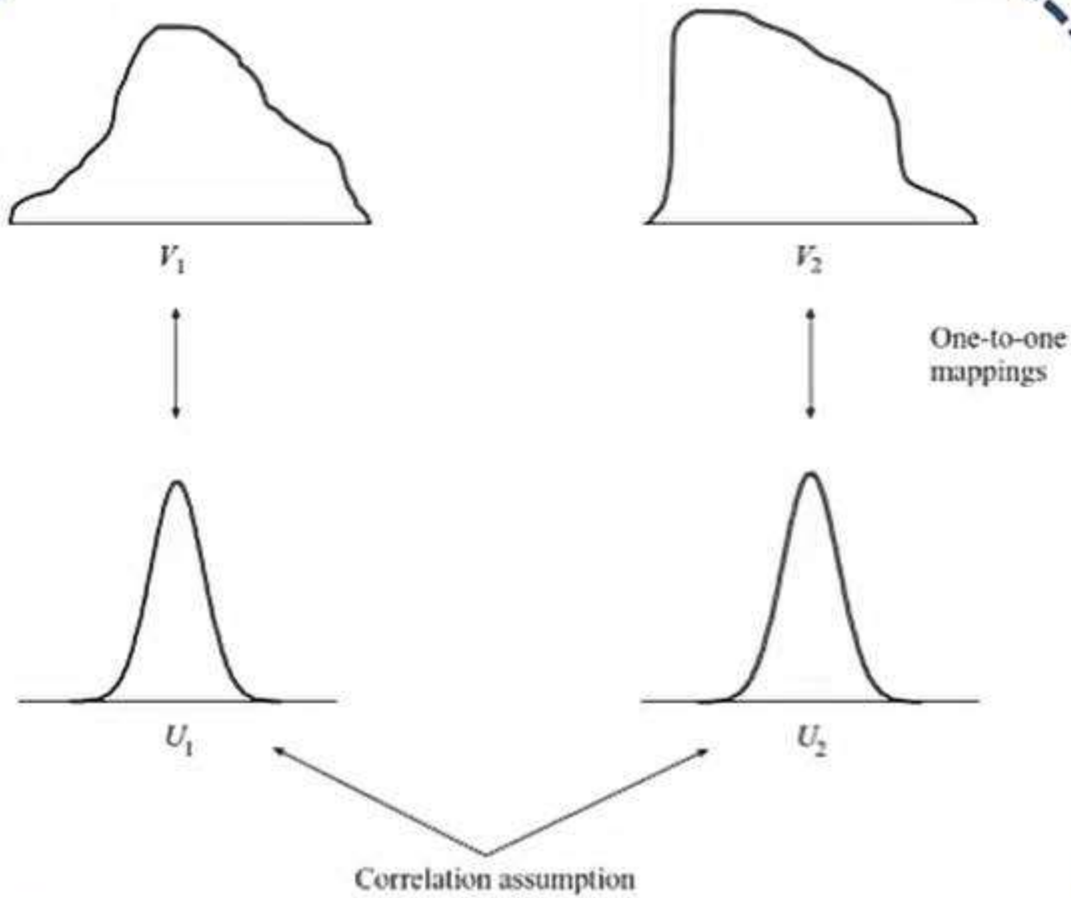


FIGURE 11.3 The Way in Which a Copula Model Defines a Joint Distribution

221-225

Copula

➢ Various Types of Copulas

- Student's t-copula: similar to Gaussian except that U_1 and U_2 are assumed to follow bivariate student's t-distribution
  - ✓ In extreme market conditions, correlations between market variables tend to increase, in that case, student's t-copula is better than Gaussian copula.
- Multivariate copula: copula can be used to define a correlation structure between more than two variables. The simplest example is the multivariate Gaussian Copula.
- One-factor (factor model) copula: $U_i = \alpha_i F + \sqrt{1 - \alpha_i^2} Z_i$

Tail Dependence

- **Tail dependence:**
  - Tail dependence is an important issue because extreme events are often related (i.e., disasters often come in pairs or more)
  - If marginal distributions are continuous, we can define a coefficient of (upper) tail dependence of X and Y as the limit, as $\alpha \to 1$ from below, of

$$\Pr\left[Y > F_y^{-1}(\alpha) | Y > F_x^{-1}(\alpha)\right] = \lambda$$

Copula

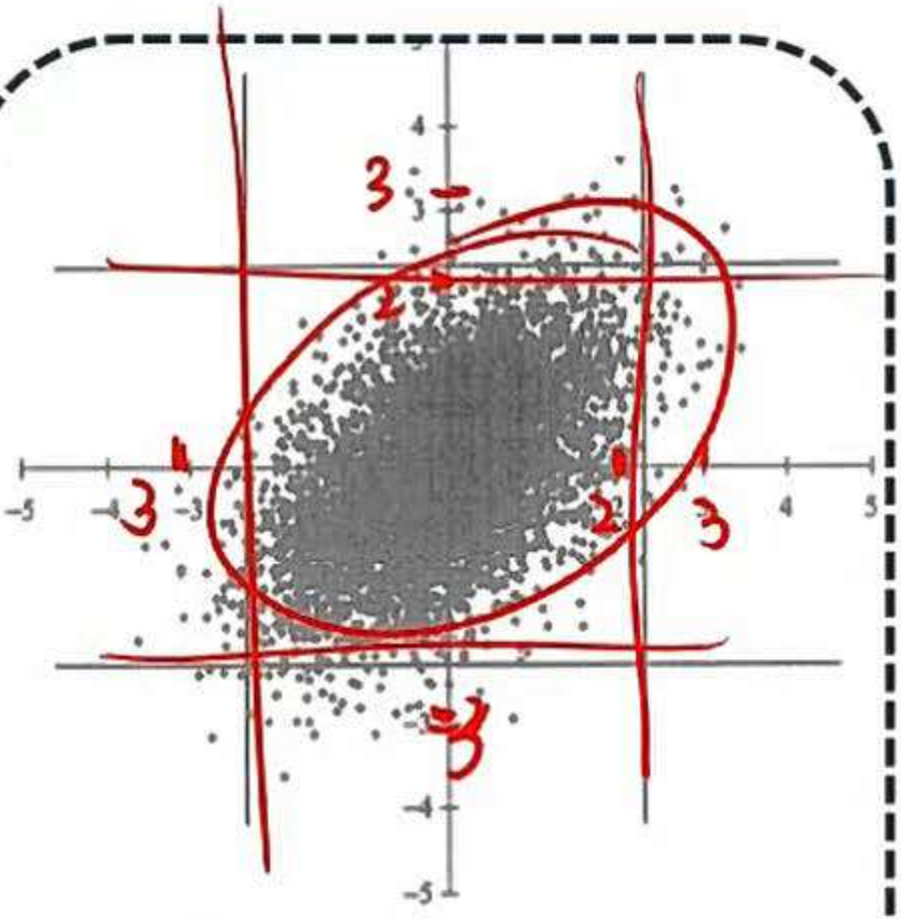


FIGURE 6-4 5,000 random samples from a bivariate normal distribution.

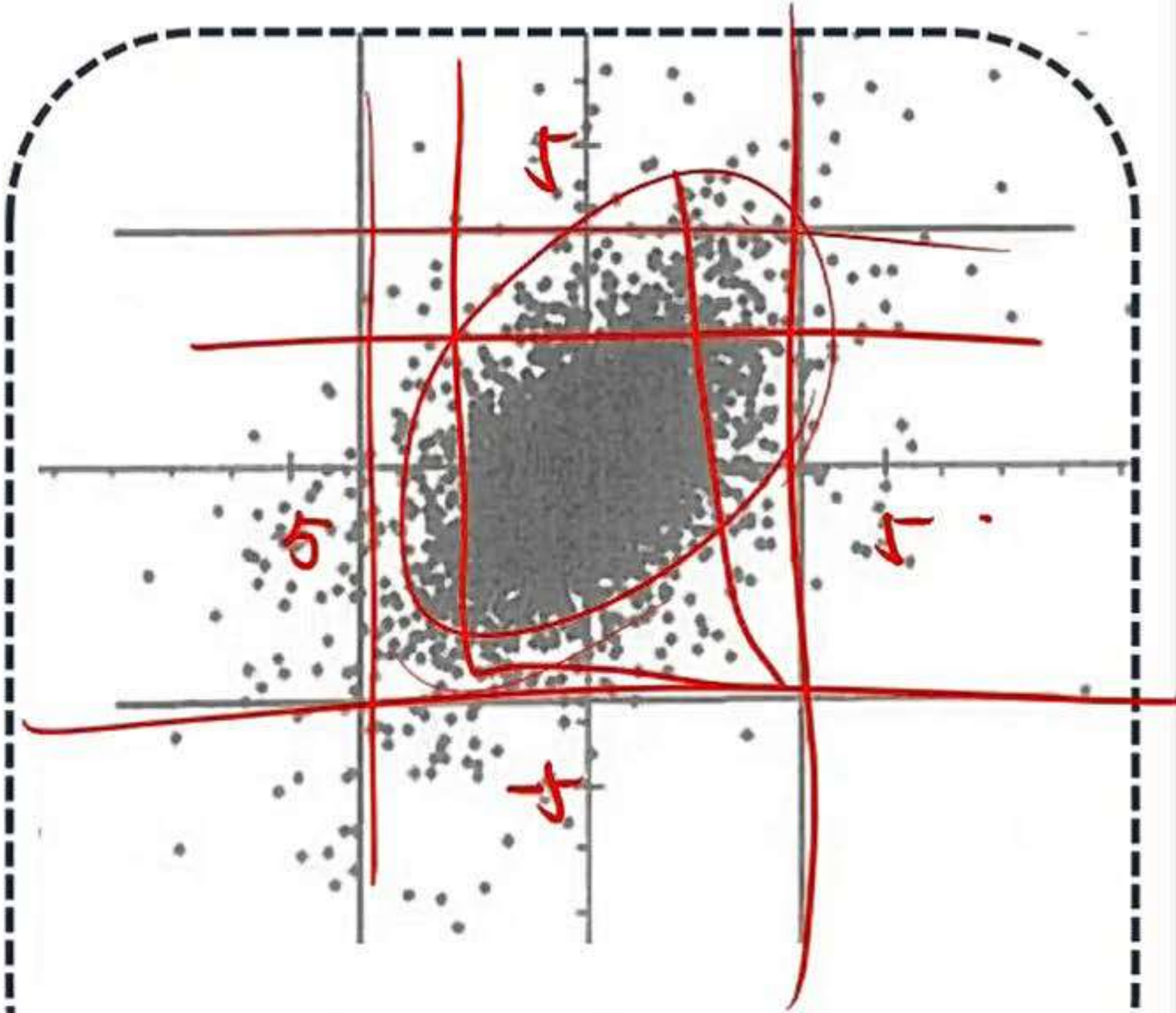


FIGURE 6-5 5,000 random samples from a bivariate Student *t*-distribution with four degrees of freedom.