CFA Pre-Courses

Quantitative Methods

Instructor: Angel

高斯

数学是“科学之王”！

不明真相的群众

研究你的数学、物理、天文去，金融跟你有什么关系！

高斯

金融里用的最广泛的概率分布“正态分布”，就是我发明的！

不明真相的群众

666666！我还以为“正态分布”是“正态”发明的呢！

笛卡儿

顶楼上上！数学是知识的工具，亦是其它知识工具的源泉！

First Insight to Quantitative Method

高顿网校 www.gaodun.com

Quant. in CFA

- Topics in CFA curriculum:

Ethics & Professional Standards	
Investment Tools	Quantitative Methods (QM)
	Economics
	Financial Reporting & Analysis
	Corporate Finance
Asset Classes	Equity Investment
	Fixed Income
	Derivatives
	Alternative Investments
Portfolio Management & Wealth Planning	

Brief Introduction

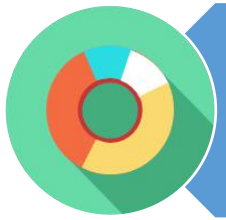
高顿网校
www.gaodun.com

学习建议:

- 能够掌握一些基本的英文单词以及翻译
- 尽量理解每一个基本概念，有利于正课更有效的学习。
- 最重要的，认真、仔细的听课。

CFA LEVEL 1: Quantitative Methods

Time Value of Money



Descriptive Statistics

Inferential Statistics

Technical Analysis

Descriptive Statistics vs. Statistical Inference

Descriptive Statistics

- Statistical Concepts
- Probability Theory

Statistical Inference

- Sampling & Estimation
- Hypothesis Test

Quantitative Methods

1 Statistical Concepts

2 Probability theory

3 Sampling & Estimation

4 Hypothesis Test

Statistical Concept

Tasks:

- Describe population, sample, parameter and sample statistics
- Identify nominal, ordinal, interval and ratio scales

Statistical Concepts

All members of a specified group.

总体
Population

样本
Sample

A subset of a population

Statistical Concepts

Research Issues:

中国金融行业的收入水平到底如何？

Statistical Concepts

Any descriptive measure of a population characteristic.

总体参数
Parameter

样本统计量
Sample Statistic

Any descriptive measure of a sample characteristic

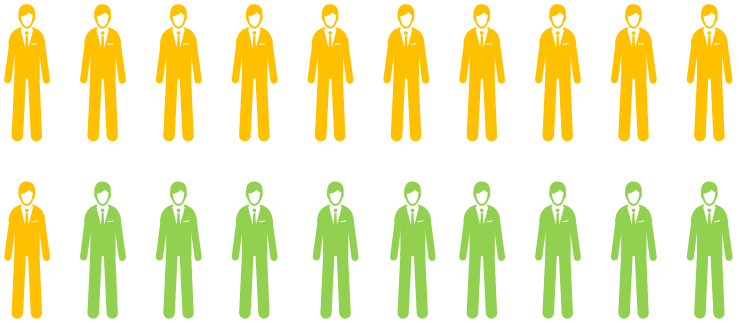
Statistical Concepts

48%

金融行业**女性**占比48%

- **平均**年收入为153,000元
- **中位数**为70,000元
- **众数**为60,000元

52%



金融行业**男性**占比52%

- **平均**年收入为155,000元
- **中位数**为72,000元
- **众数**为68,000元

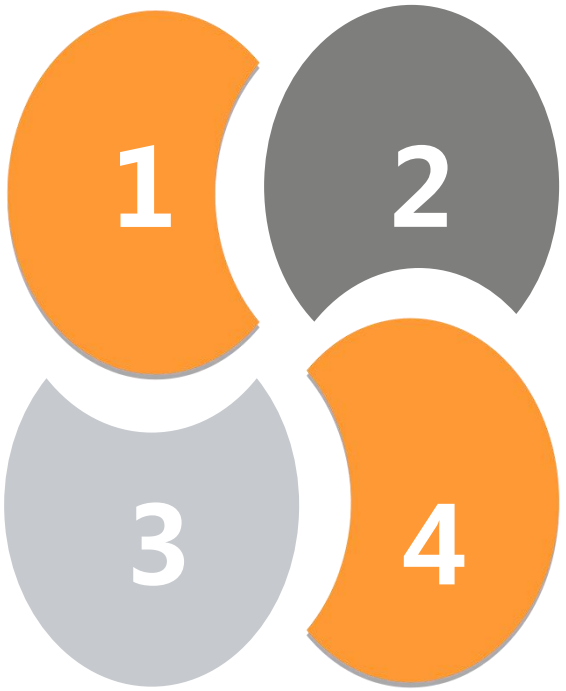
Measurement Scales

Nominal Scales

- Category but no ranking
- **Weakest** level of measurement

Ordinal Scales

- Sorting data into categories that are ordered with respect to some characteristics



Interval Scales

- Ranking the data with assurance that the differences between scale values are equal

Ratio Scales

- Having all the characteristics of interval measurement scales as well as true zero point as the origin;
- **Strongest** level of measurement

Measurement Scales

高顿网校 www.gaodun.com

- Hedge Fund Styles – **Nominal Scales**

George Scoros

全球最大的投资者，与“商品大王”吉姆罗杰斯（Jim Rogers）合伙成立了“量子基金”

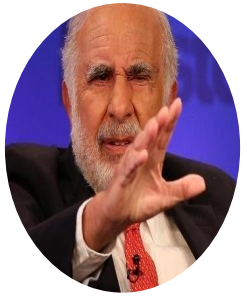
John Paulson

Paulson & Co公司总裁，因在2008年美国次贷危机中大肆做空而获利，被认称为“华尔街空神”

David Tepper

对冲基金Appaloosa创始人。“抄底王，冷门投资大师，典型的机会主义者”

Carl Icahn

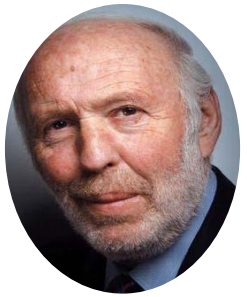


75岁的他仍然无退休之意，爱股市爱期货，最爱恶意收购。被称为CEO公敌，华尔街之狼。

Steve Cohen

SAC Capital Advisors(SAC资本)创始人，管理者140亿美元的投资组合，主要做短线交易

James Simons



全球收入最高的对冲基金经理，文艺复兴科技公司荣誉主席职位

Measurement Scales

高顿网校 www.gaodun.com

- Private Equity Ranking in China – **Ordinal Scales**

1 KKR

KKR

老牌的杠杆收购天王，金融史上最成功的产业投资机构之一

2

阿里资本

通过投资、并购和业务拓展，创造战略和长远财富价值

3

晨晖创投

信息技术领域的早期投资

从容投资

专注于价值投资

4

淡马锡投资

多元化投资集团

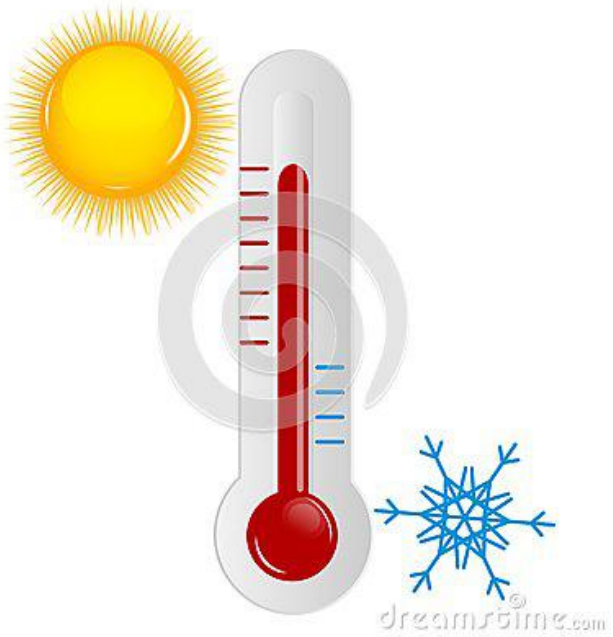
TEMASEK 5

鼎锋资产

专注于成长股投资

6

Measurement Scales



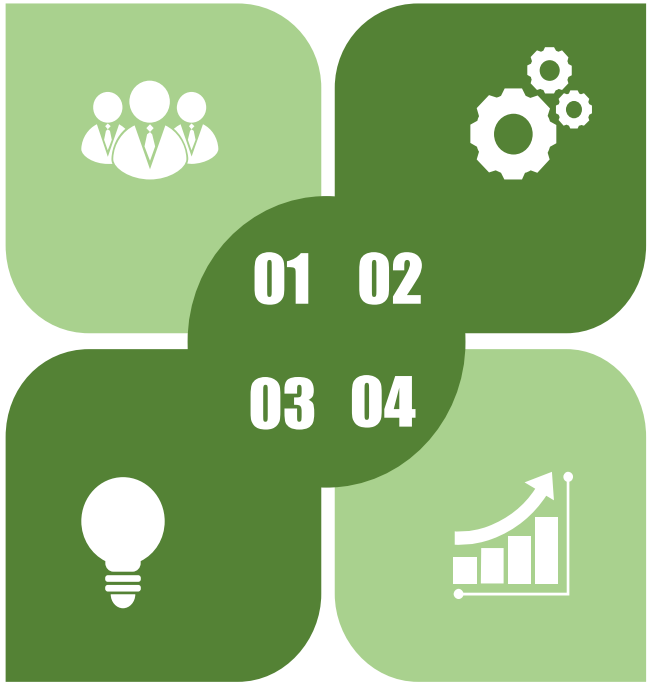
Temperature
Interval Scales

Wealth
Ratio Scales

Practice Question

Question：Identifying Scales of Measurement

Credit Ratings for bond issues



Cash dividends per share

Hedge fund classification types

Bond maturity in years

Practice Question

Answer:

1. **Ordinal scale**; A rating places a bond issue in a category, and the v=categories are ordered with respect to the expected probability of default. But the difference between AA-and A+ is not necessarily squal to that between BB- and B+
2. **Ratio scale**; 0 represents the complete absence of dividends, it is a true zero point.
3. **Nominal scale**; Hedge fund classification schemes do not involve a ranking.
4. **Ratio scale**.

Central Tendency

Tasks:

- Calculate and Interpret Arithmetic mean, geometric mean, weighted mean and Harmonic mean.
- Describe mode and median.
- Calculate and interpret quantile

Quantitative Description of Distribution

Description

1 Central Tendency

2 Quantile

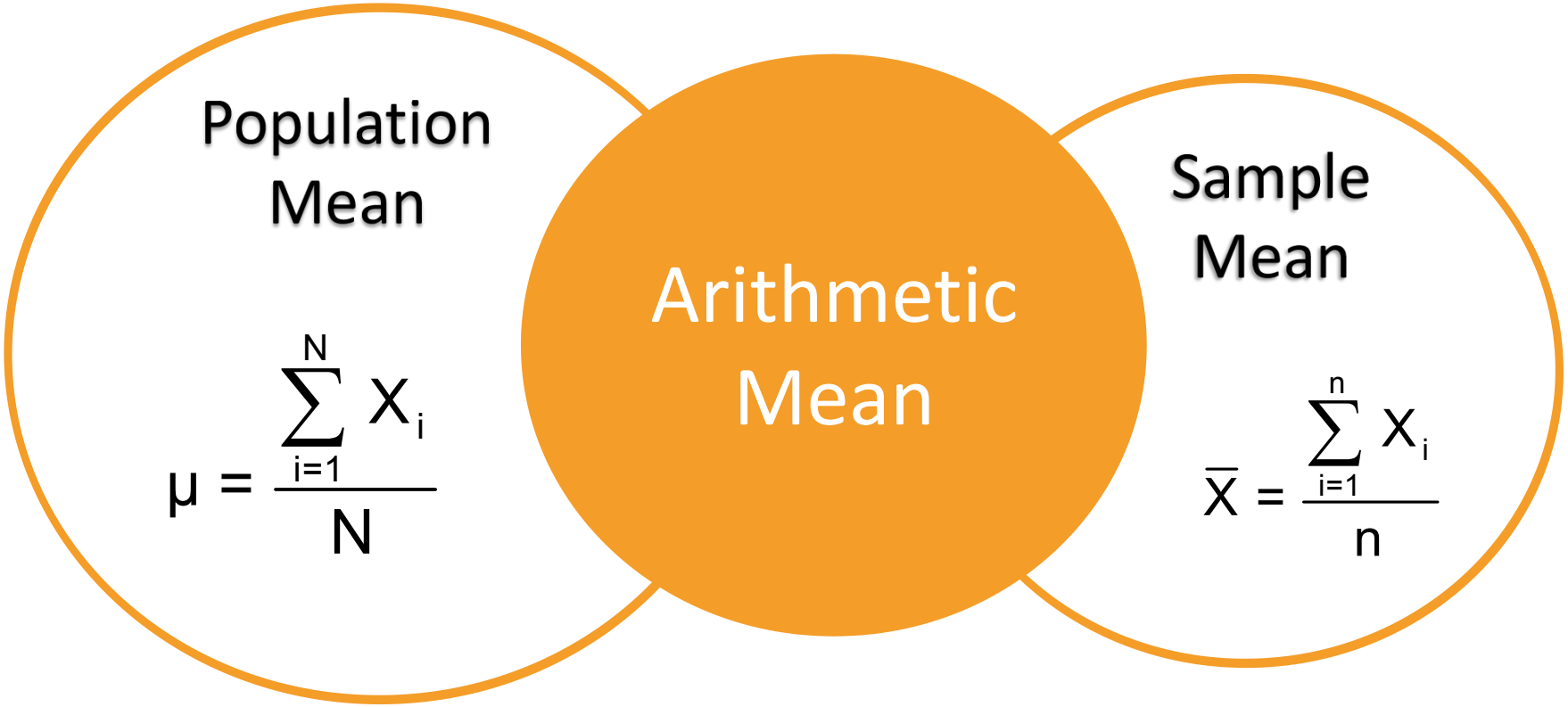
3 Dispersion

Central Tendency

统计数据显示，2007年至2014年定增**平均收益率达56%**，其中，2014年参与一级市场的**定增**平均收益率为80%，**约96%的项目可获绝对收益。。。**

Central Tendency

- Arithmetic mean
  - The sum of the observations divided by the number of observations



过去三年，我的第一年收益率100%，第二年收益率100%，第三年收益率-100%，所以平均年收益率是33%么？

实际情况是连本金都没有了！

高顿网校
www.gaodun.com

Central Tendency

- Geometric Mean
  - It is widely used to compute the compound growth rate of a variable

$$G = \sqrt[n]{X_1 X_2 X_3 \ldots X_n} \qquad \text{with} \quad X_i \geq 0 \text{ for } i = 1, 2, 3, \ldots, n.$$

$$\text{Periodic return}_{\text{compound}} = \sqrt[n]{(1+R_1)(1+R_2)\ldots\ldots\ldots(1+R_n)} - 1$$

  - Arithmetic mean is always greater than or equal to the Geometric Mean.

Central Tendency

- Harmonic Mean
  - It is widely used to compute the average cost for purchasing

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

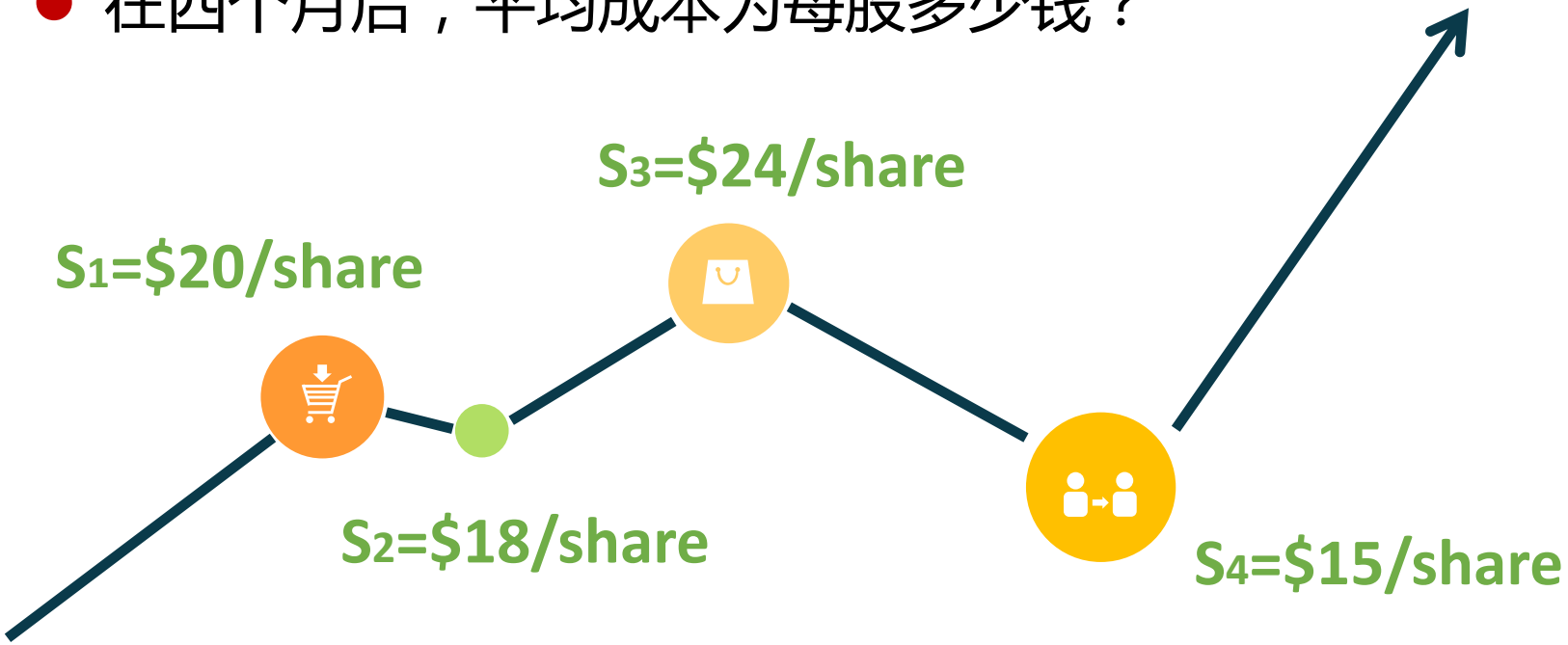
where: N = number of purchases (equal $ amounts)

X_i = share price for each purchase

  - TIPS：调和平均的倒数就是倒数的平均

Central Tendency

- 基金定投：某基金每月固定投资30万美元于某股票
  - 在四个月后，平均成本为每股多少钱？



  - Average cost per share

    =4/(1/20+1/24+1/18+1/15)=$18.70/share

Central Tendency

- Weighted Mean
  - An average in which each observation is weighted by an index of its relative importance.

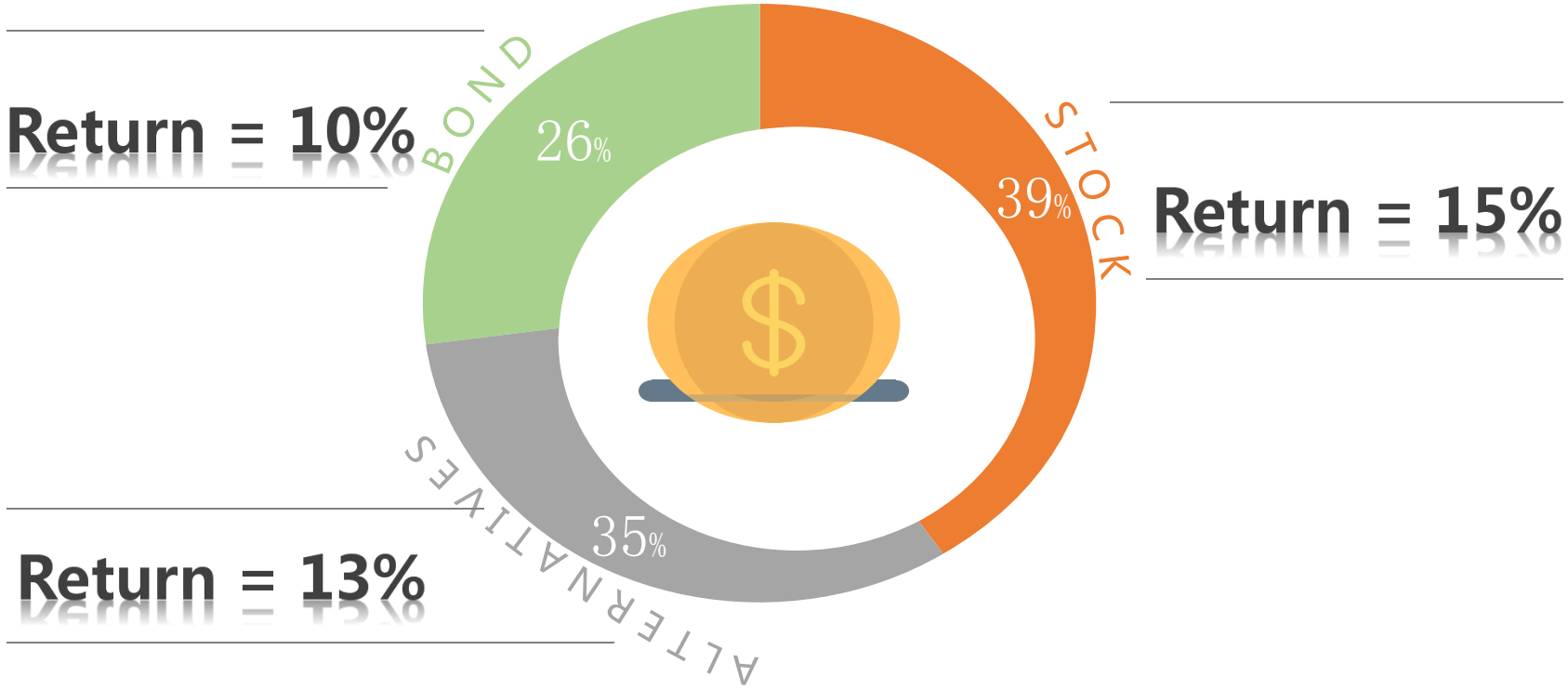
$$\overline{X}_w = \sum_{i=1}^{n} w_i X_i$$

Where the sum of the weights equals 1;

that is, $\sum_i w_i = 1$

Practice Question

- 某基金经理给其客户分别配置了股票，债券和其他类的产品
  - 该客户当年的回报是多少？

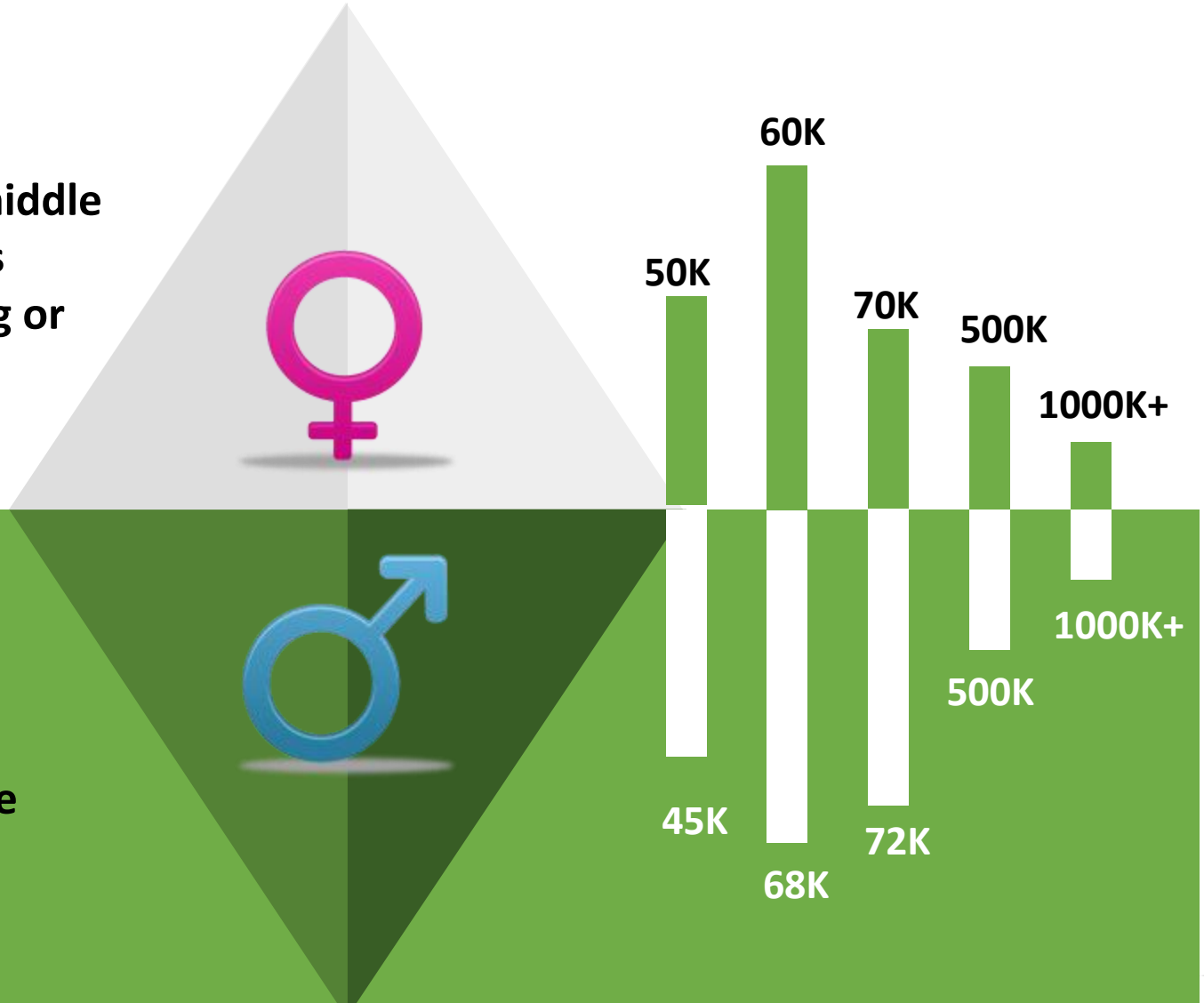


  - Total Return = 26%*10%+39%*15%+35%*13% = 13%

Central Tendency

Median

➢ The value of the middle item of a set of items sorted into ascending or descending order.



Mode

➢ Most Frequently occurring value of the distribution

Practice Question

- Suppose a client asks you for a valuation analysis on the seven-stock US common stock portfolio given in the table below.

Stock	Consensus Current EPS	Consensus Current P/E
Catepillar, Inc (NYSE: CAT)	6.34	13.15
Ford Motor Company (NYSE: F)	1.55	10.97
General Dynamics (NYSE:GD)	6.96	12.15
Green Mountain Coffee Roasters (NASDAQ: GMCR)	3.25	25.27
McDonald's Corporation(NYSE: MCD)	5.61	17.16
Qlik Technologies (NASDAQ: QLIK)	0.17	204.82
Questcor Pharmaceuticals (NASDAQ: QCOR)	4.79	13.94

Practice Question

Question 1: Calculate the arithmetic mean P/E

The mean P/E is

(13.15+10.97+12.15+25.27+17.16+204.83+13,94)/7=42.49

Question 2: Calculate the median P/E

The P/Es listed in ascending order are:

10.97 12.15 13.15 13.94 17.16 25.27 204.82

The median P/E is 13.94

Quantile

- **Quantile**: A value at or below which a stated fraction of the data lies.

Quartiles
The distribution divided into quarters

Quintiles
The distribution divided into the fifths

Deciles
The distribution divided into the tenths.

Percentiles
The distribution divided into the hundredths

$$L_y = (n+1)\frac{y}{100}$$

Where: n=the number of data

y=the y^{th} percentile

Practice Question

- The EURO STOXX 50 is an index of 50 publicly traded companies. The table shows the market capitalization on the 50 component stocks in the index, as provided by STOXX Ltd. In September 2013. The market capitalization are ranked in ascending order.

No.	Company	Market Cap (Euro Billion)	No.	Company	Market Cap (Euro Billion)
1	Arcelor-Mittal	8.83	37	Telefonica	39.00
2	CRH	10.99	38	ENI	41.42
3	RWE	11.92	39	Daimler	42.42
4	Carrefour	12.13	40…	BNP Paribas	43.09
5	Repsol	12.84	48	Bayer	65.83
6	Saint-Gobain	13.60	49	Total	81.06
7…	France Telecom	14.09	50	Sanofi	93.29

Practice Question

Q1. Calculate the 10th percentiles

$L_{10} = (50+1)(10/100)$
$= 5.1$

L_{10} is between the fifth and sixth observation with value $X_5 = 12.84$ and $X_6 = 13.6$.

$P_{10} \approx X_5+(5.1-5)(X_6-X_5)$
$= 12.84 + 0.1(13.60-12.84)$
$= 12.92$

Q2. Calculate the third quartiles.

$L_{75} = (50+1)(75/100)$
$= 38.25$

L_{75} is between the 38th and 39th observation with value $X_{38} = 41.42$ and $X_{39} = 42.42$.

$P_{75} \approx X_{38}+(38.25 - 38)(X_{39}-X_{38})$
$= 41.42 + 0.25(42.42-41.42)$
$= 41.67$

Dispersion

Tasks:

- Calculate and Interpret Range, Mean Absolute Deviation, Variance and Standard Deviation

Dispersion

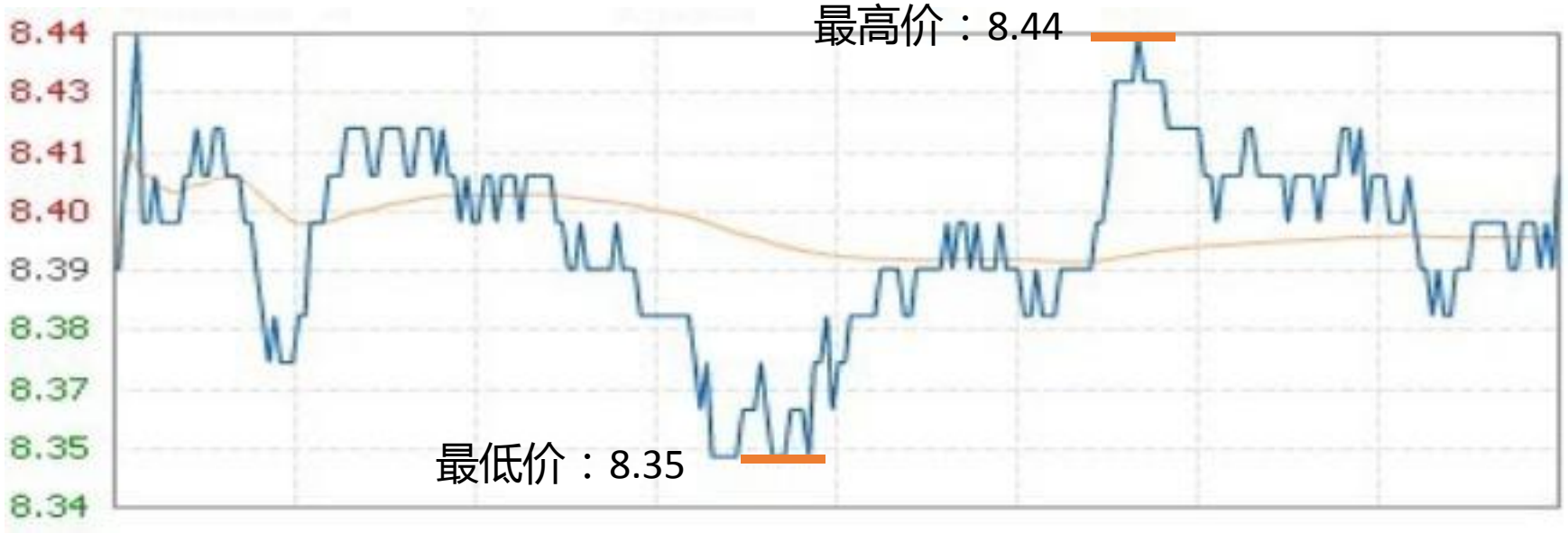
“这事儿没啥风险”

“不知道风险存在就是最大的风险”

“风险的大小取决于不确定性的复杂程度，时间长短，目标高低以及可承受能力的强弱”

从量化的角度来衡量风险

Dispersion



- Range = Maximum Value - Minimum Value = 8.44–8.35 = 0.09
- Mean Absolute Deviation (MAD)

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

Where : $\bar{X}$ is the sample mean, n is the number of observations.

Practice Question

Year	Selected American Shares (SLASX)
2008	-39.44%
2009	31.64
2010	12.53
2011	-4.35
2012	12.82

Q1: Calculate the range of annual returns for SLASX

For SLASX, the largest return was 31.64 percent and the smallest was -39.44 percent. The range is thus 31.64 – (-39.44)=71.08%

Q2: Calculate the mean absolute deviation of returns on SLASX.

The mean return for SLASX is 2.64 percent.

$$MAD=\frac{|-39.44-2.64|+|31.64-2.64|+|12.53-2.64|+|-4.35-2.64|+|12.82-2.64|}{5}=\frac{98.14}{5}=19.63\%$$

Dispersion

- **Variance**: equal to average of the sum of squared deviations around the mean.

Population Variance

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

Sample Variance

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

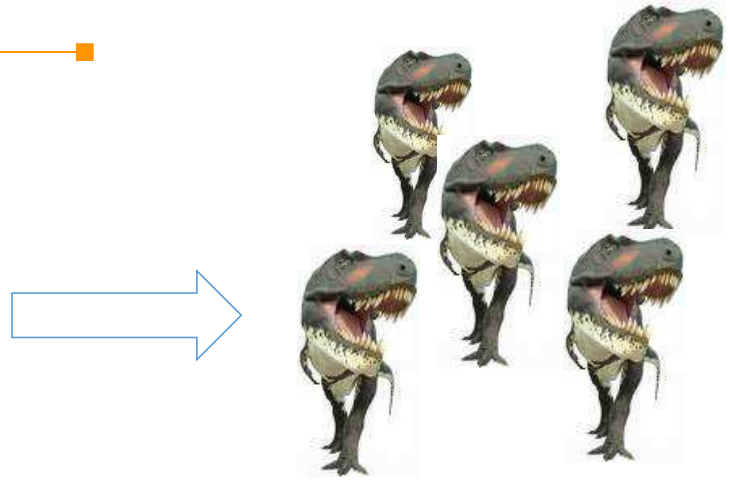
- **Standard Deviation**：positive squared root of variance.

Probability Concepts

Tasks:

- Describe the basic probability concepts
- Calculate probability using addition rule, multiplication rule

Probability VS statistics



Statistic

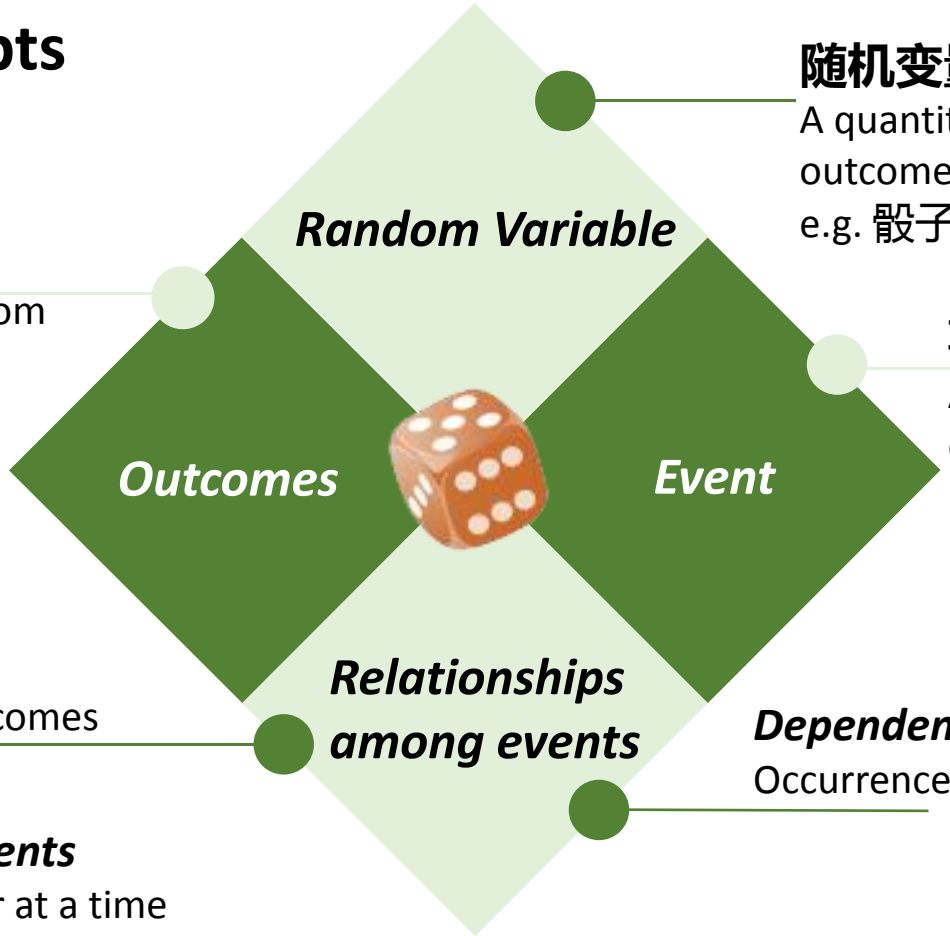
We have a sample, and infer about an unknown population

Probability

We can known the population, and ask the probability of getting the sample.

Probability Concepts

Basic concepts



随机变量（研究对象）
A quantity whose future outcomes are uncertain
e.g. 骰子扔下去的点数, X

随机变量可能的取值
Possible values of a random variable
e.g. X=1, X=2, X=3

事件
A specified set of outcomes
e.g. X=1, X=even, X>2

Exhaustive events
Contains all possible outcomes
e.g. X = 1,2,3,4,5,6

Mutually exclusive events
Only one event can occur at a time
e.g. X=2与X=3互斥

Dependent events
Occurrence of A is related to B

Probability Concepts

1. Probability of stock earns a return above risk-free rate.
2. Probability of stock earns a return above risk-free rate, given that the stock earns a positive return.

事件1，2 的区别是什么？

- ***Unconditional probability/marginal probability***
  - P(A) 事件A发生的概率
- ***Conditional probability***
  - $P(A|B) = \frac{P(AB)}{P(B)}$ 事件B发生的前提下，事件A发生的概率
- ***Joint Probability***
  - P(AB) 事件A与B同时发生的概率

Probability Concepts

- ***Multiplication rule,*** which is conditional probability rearrange.

$$P(AB) = P(A|B)P(B)$$

- ***Addition rule***

$$P(AorB) = P(A) + P(B) - P(AB)$$

如果事件A和B是独立事件？

$$P(A|B) = P(A)$$

如果事件A和B是互斥事件？

$$P(AB) = 0$$

Practice Question

- Funds are classified as benchmark-adjusted return for each of two consecutive years. The top 50 percent of funds by benchmark-adjusted return for a given year are labeled winners; the bottom 50 percent were labeled losers.
- *For example*, winner-winner entry shows that 65.5% of the first-year winner funds were also winners in the second year. Note that the four entries in the table can be viewed as conditional probabilities.

	Year 2 Winner	Year 2 Loser
Year 1 winner	65.5%	34.5%
Year 1 loser	15.5%	84.5%

Practice Question

Q1: State the four entries of the table as conditional probability using the form P(this event | that event) = number

From Row 1:

P(fund is a Year 2 winner|fund is a Year 1 winner) = 0.655

P(fund is a Year 2 loser | fund is a Year 1 winner)=0.345

From Row 2:

P(fund is Year 2 winner |fund is a Year 1 loser) = 0.155

P(fund is a Year 2 loser | fund is a Year 1 loser) = 0.845

Practice Question

Q2: Using information in the table, calculate the probability of the *event a fund is a loser in both Year 1 and Year 2*.

Note that because 50 percent of funds are categorized as loser in each year, the unconditional probability that a fund is labeled a loser in either year is 0.5.

From the table, $P(A|B)=0.845$ and $P(B)=0.5$.

$P(AB) = P(A|B)P(B)=0.845(0.50)=0.4225$

Bayes' Formula

Tasks:

- Calculate and interpret Bayes' probability
- Distinguish covariance and correlation coefficient

Probability Concepts

- ***Total probability***

$$P(A) = P(A|S_1) \times P(S_1) + P(A|S_2) \times P(S_2) + P(A|S_n) \times P(S_n)$$

  - ✓ where S1 S2…Sn are mutually exclusive and exhaustive

- ***Bayes' formula***

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

  - ✓ P(A)反映了过去的已掌握的信息，P(A|B)则反映了事件B发生后对P(A)的修正。

Practice Question

A上市公司即将要公布其上季度财报。作为一名投资者，你对上季度EPS能否超越 consensus EPS estimate（市场预估EPS）很感兴趣。

基于自己的研究，你给出以下概率：

P(EPS exceeded consensus) = 0.45

P(EPS met consensus) = 0.30

P(EPS fell short of consensus) = 0.25

第二天A公司公布将在新加坡扩大产能的消息, based on the new information, what's the **updated probability** that prior quarter EPS exceeded consensus EPS?

Practice Question

Suppose you know:

P(A expands | EPS exceeded consensus) = 0.75

P(A expands | EPS met consensus) = 0.20

P(A expands | EPS fell short of consensus) = 0.05

Answer:

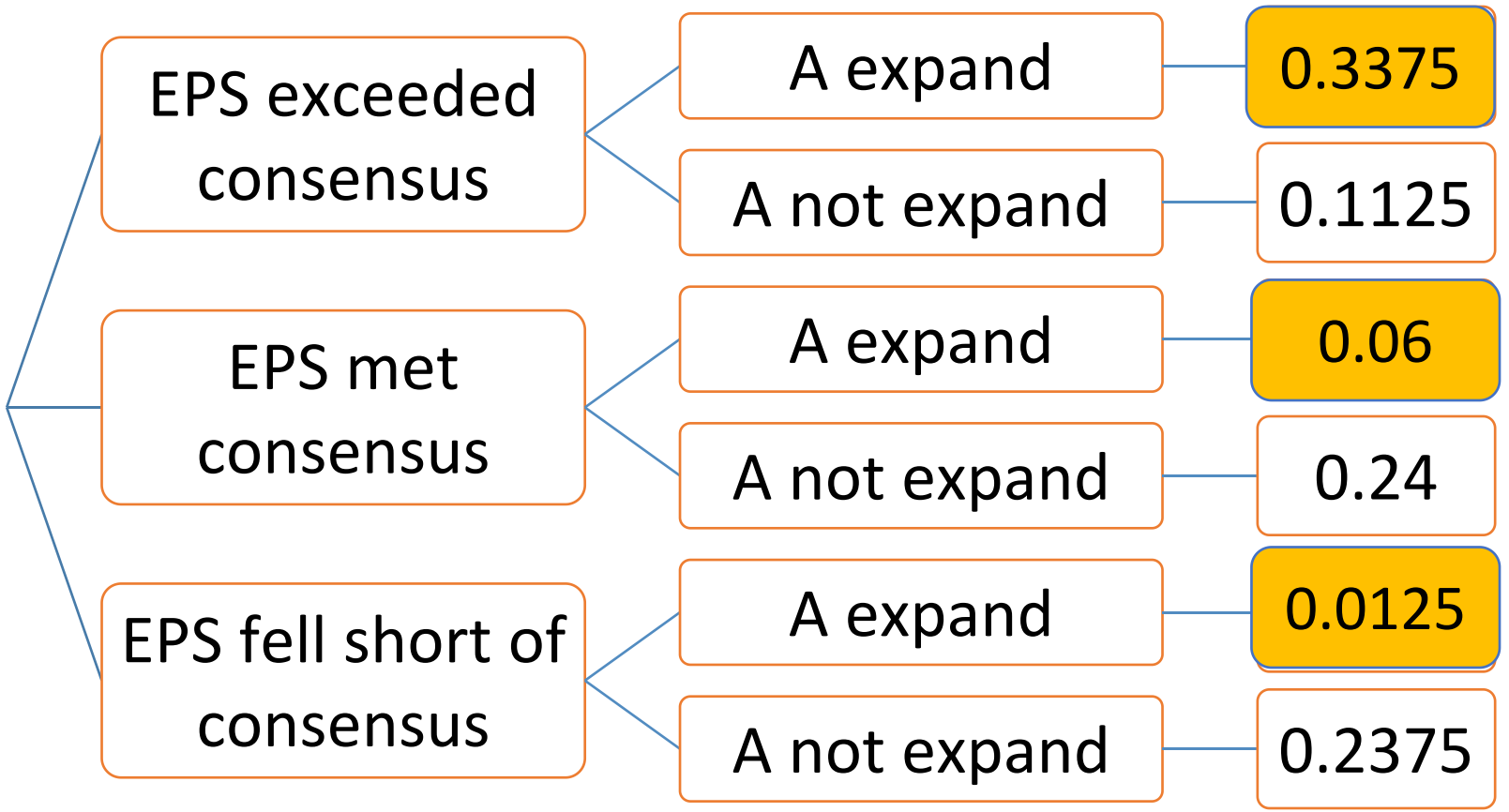
P(EPS exceeded consensus |A expands)

=[P(A expands| EPS exceeded consensus)/P(A expands)]

X P(EPS exceeded consensus) = 0.75/0.41*0.45 = 0.823

Which, P(A expands)= 0.45*0.75+0.3*0.2+0.25*0.05=0.41

Practice Question



P(EPS exceeded consensus |A expands)
= (0.3375)/(0.3375+0.06+0.0125)=0.823

Probability Concepts

高顿网校 www.gaodun.com

Covariance	VS	Correlation
$Cov(R_i, R_j) = E[(R_i - ER_i)(R_j - ER_j)]$		
The **co-movements** of 2 random variables		**Linear relationship** of 2 random variables
协方差		**相关性**

Covariance =0 两个资产的收益率之间没有线性关系

Covariance >0 两个资产的收益率同时大于或者同时小于它们各自的期望值

Covariance <0 一个资产的收益率大于其期望而另一个的小于其期望值

$$-1 \le \rho(R_i, R_j) \le 1$$

ρ=0 两个资产的收益率不存在线性关系

ρ=1 perfect linear relationship

ρ=-1 perfect inverse linear relationship

Discrete Distribution

Tasks:

- Describe discrete and continuous distribution
- Calculate the probability on binomial random variable

Probability Distribution

- It specifies probability of all possible outcomes for a random variable.

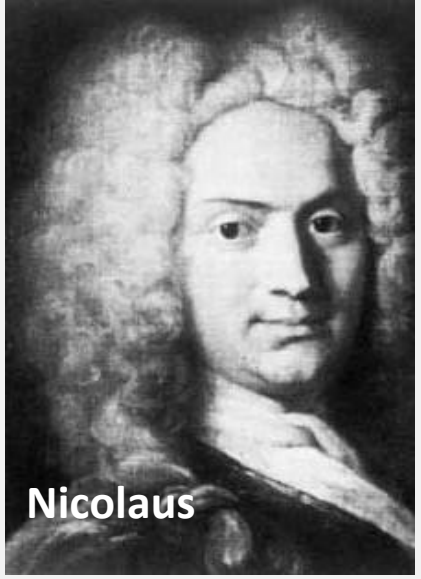
Probability Distribution

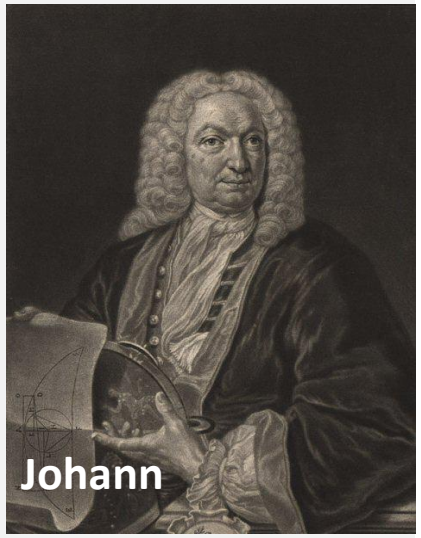
Discrete variable

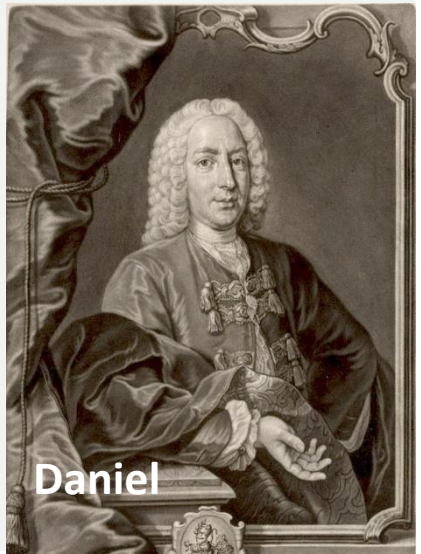
Discrete distribution

Continuous variable

Continuous distribution







这里盛产数学家和物理学家！

他们是数学史上的奇迹！

他们是一个神话!

雅各布•伯努利 Jakob Bernoulli

被公认的概率论的先驱之一。他是最早使用“积分”这个术语的人，也是较早使用极坐标系的数学家之一。还较早阐明随着试验次数的增加，频率稳定在概率附近。概率论中的**伯努利试验**与大数定理也是他提出来的。

Probability Distribution

Bernoulli Trial

- 伯努利试验是“在同样条件下进行重复试验”的一种数学模型。
- 在一次伯努利试验中，事件A发生的概率为p，不发生的概率为1-p。即 P(X=1) = p，P(X=0) = 1-p

Probability Distribution

Binomial random variable

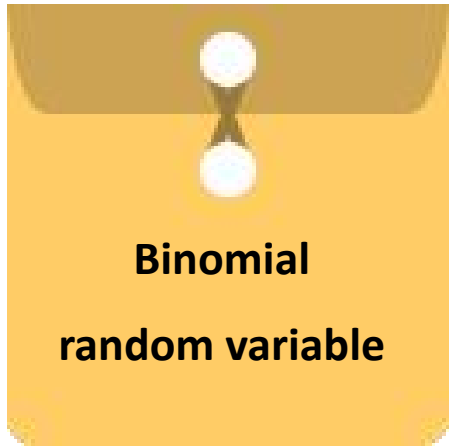
- 定义：n次伯努利试验中成功的次数。
- 前提：
  - 每一次试验都是独立的；
  - 每一次成功的概率均相等为p
- 二项随机变量的概率：

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

Practice Question

An investment newsletter claims a 70 percent success rate in making investment recommendations that are profitable. You have the newsletter's seven most recent recommendations.

All the recommendations are independent and the newsletter writer's skill is as claimed.



QUESTION:

what is the probability of observing two or fewer profitable recommendations out of the total?

Practice Question

Answer:

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

The probability of two or fewer successes is

$F(2) = p(2) + p(1) + p(0)$ $n = 7$ and $p = 0.70$

$p(2) = (7!/2!5!)(0.70)^2(0.30)^5 = 0.025005$

$p(1) = (7!/1!6!)(0.70)^1(0.30)^6 = 0.003572$

$p(0) = (7!/0!7!)(0.70)^0(0.30)^7 = 0.000219$

$F(2) = 0.025005 + 0.003572 + 0.000219 = 0.028796$

Continuous Distribution

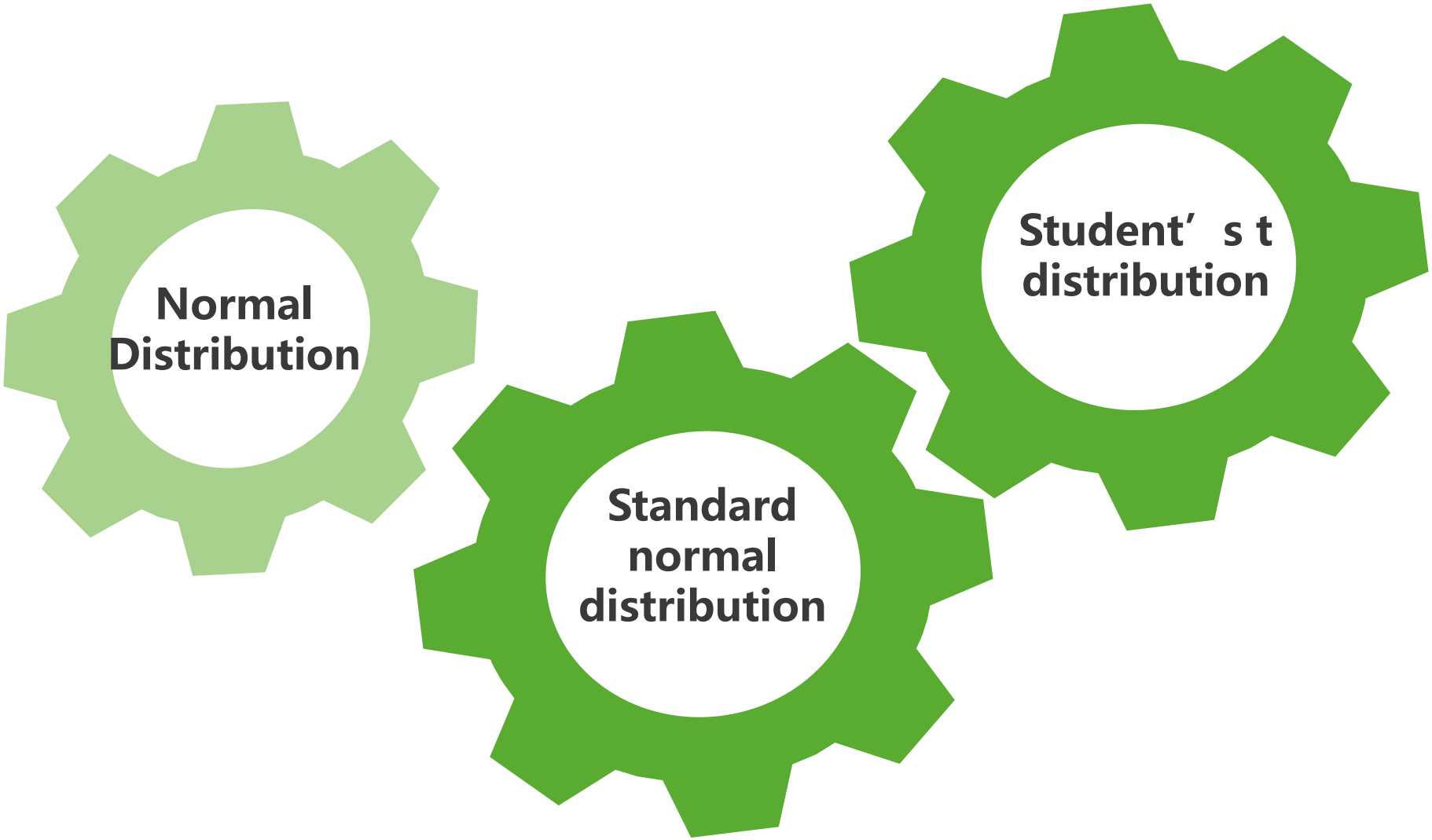
Tasks:

- Describe and interpret the normal distribution
- Calculate and interpret z value
- Distinguish the normal distribution and student's t distribution

Probability Distribution

高顿网校 www.gaodun.com

Continuous distributions



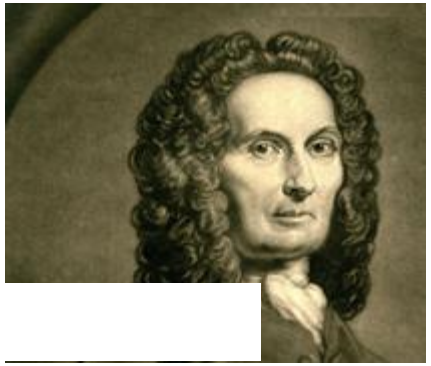
1667-1754
法国数学家

1733年

De Moivre在研究二项分布的随机变量的过程中，第一次发现了正态分布的概率分布曲线。

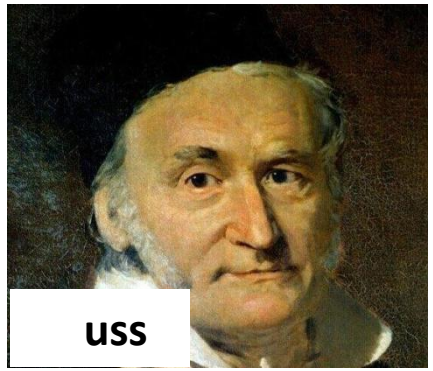
1770年

Laplace将二项分布的正态近似推广到任意p的情况，并以二项分布的极限形式推导出了正态分布的密度函数。



1749-1827
法国物理学家、数学家，有“法国的牛顿”之称。

De Moivre-Laplace中心极限定理



1809年

正态分布也被称为“高斯分布”
Gauss证明了最小二乘法拟合过程中产生的随机误差服从正态分布。在高斯的推动下，正态分布迅速在测量误差分析中被广泛使用。

1777-1855
德国数学家、物理学家、天文学家、大地测量学家。是近代数学奠基者之一，高斯和阿基米德、牛顿并列为世界三大数学家。

Probability Distribution

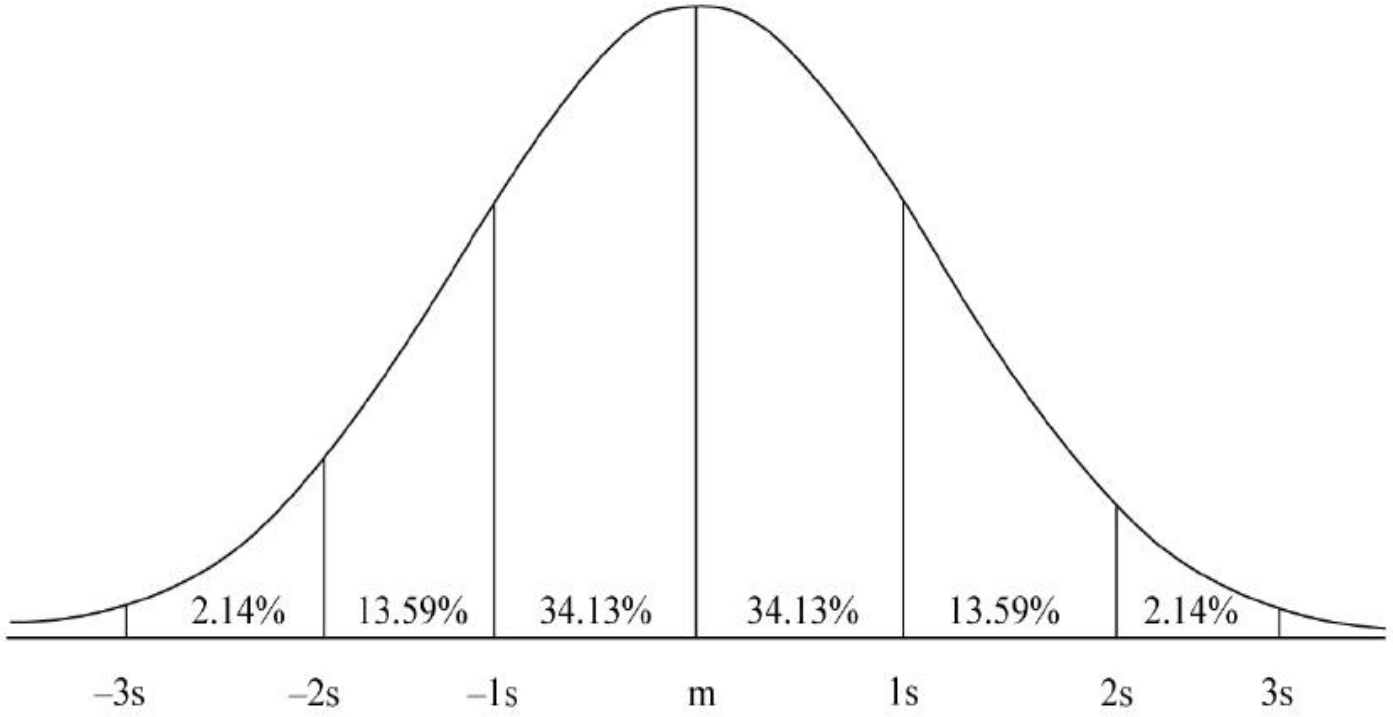
Normal Distributions

- 其随机变量X的概率密度为：

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \text{ for } -\infty < x < +\infty$$

- 其中μ和 σ^2 都是常数，μ任意， σ > 0，则称X服从参数为μ和 σ^2 的正态分布。
  - 记作 X ~ N(μ, σ^2)
- 正态分布的密度曲线是一条关于μ对称的钟形曲线。特点是“两头小，中间大，左右对称”。

Probability Distribution

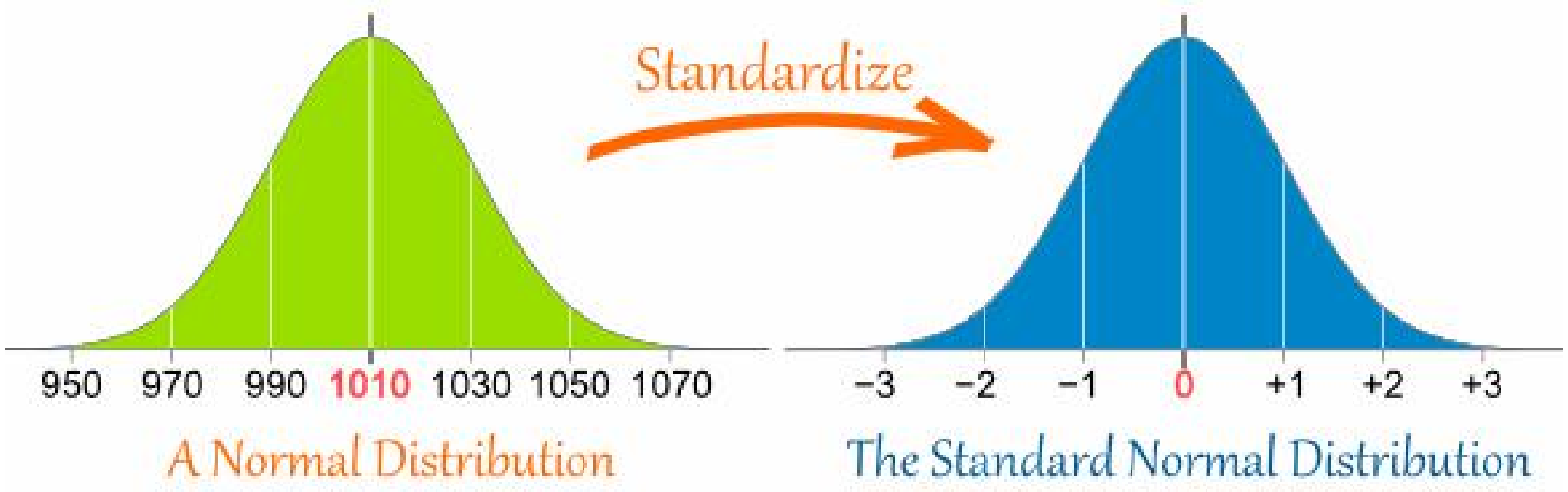


Approximately 68% of all observations fall in the interval (μ+/- σ)
Approximately 95% of all observations fall in the interval (μ+/- 1.96σ)
Approximately 99% of all observations fall in the interval (μ+/- 2.58σ)

Probability Distribution

Standard normal distributions（标准正态分布）

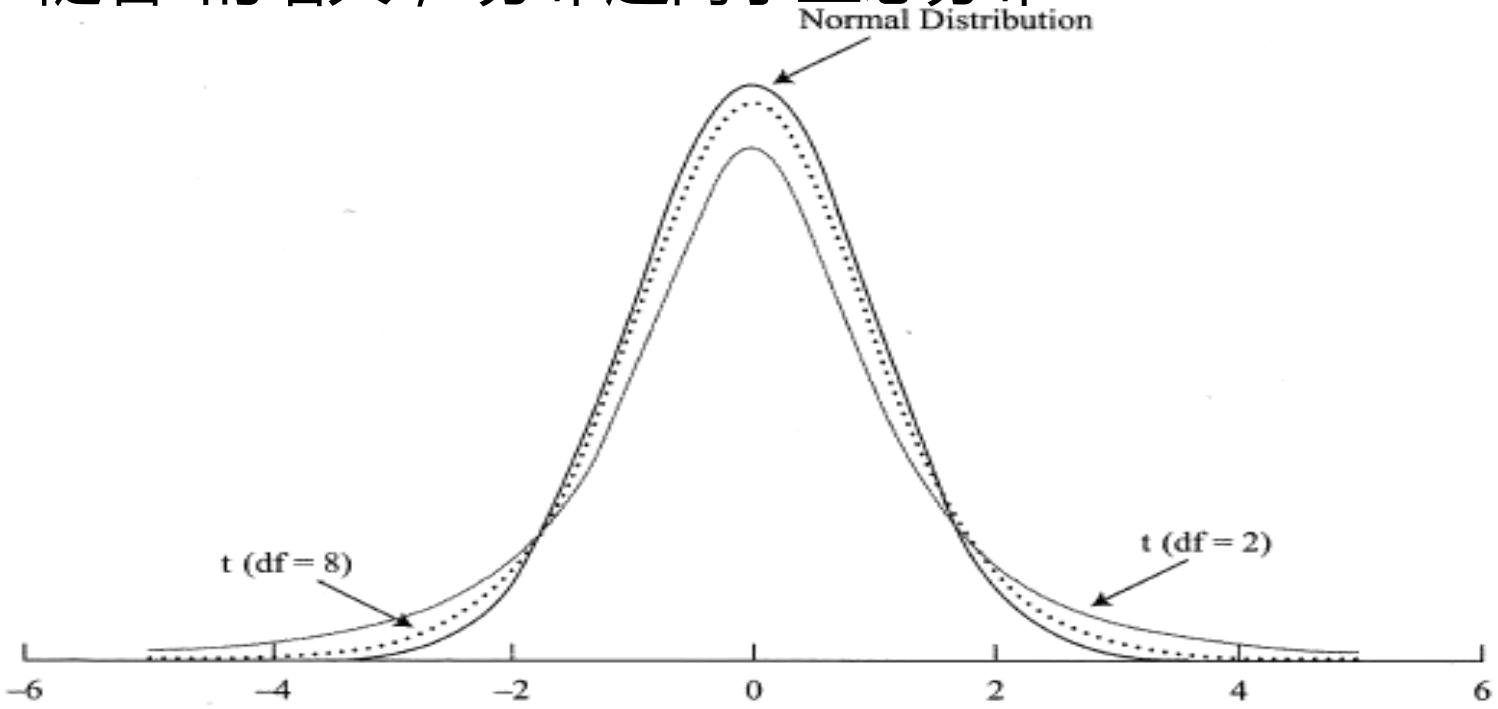
- 若X ~ N(μ, σ^2），则可以通过 Z=（X-μ）/σ 将其标准化
- Z ~N（0，1）



Probability Distribution

Student's t distribution

- 尾部比正态分布宽，自由度n-1, 样本量(n)不同，t分布也不同
- 随着n的增大，t分布趋向于正态分布

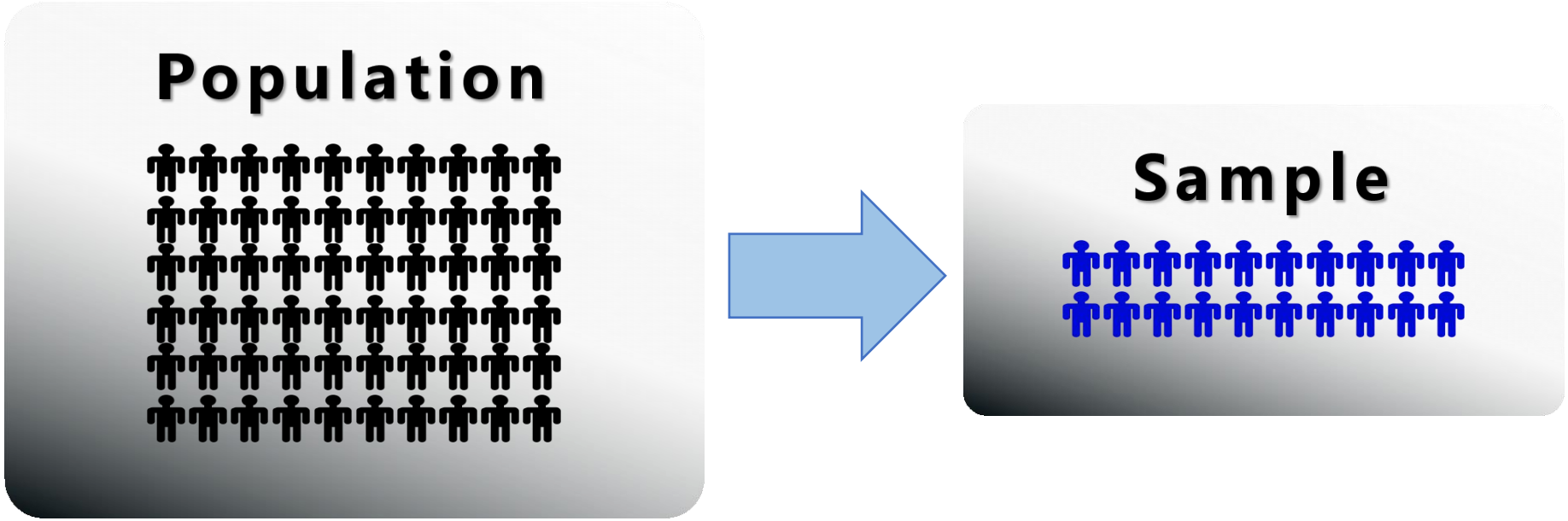


Sampling

Tasks:

- Clarify time series data and cross sectional data
- Describe the sampling process
- Describe and interpret the sampling bias
- Explain sampling distribution and sampling error

Sampling



Why Sampling?

1. We cannot possibly examine every member of the population.
2. Examining every member of the population would not be economically efficient.

Sampling

Time-Series Data

Cross Sectional Data

US Dollar	1.00 USD	inv. 1.00 USD
Euro	0.902802	1.107662
British Pound	0.754532	1.325324
Indian Rupee	67.518482	0.014811
Australian Dollar	1.343285	0.744444
Canadian Dollar	1.299170	0.769722
Singapore Dollar	1.347940	0.741873
Swiss Franc	0.977053	1.023485
Malaysian Ringgit	3.994851	0.250322
Japanese Yen	102.710327	0.009736
Chinese Yuan Renminbi	6.648746	0.150404

Sampling

Simple Random sampling

- A *subset* of a larger population created in such a way that each element of the population has an *equal probability* of being selected to the subset.

Simple Random sampling

- 如何使用简单随机抽样的方法进行电话访问获取金融行业的真实收入水平？

If the *population* contains 500,000 members, we number them from *1 to 500,000*

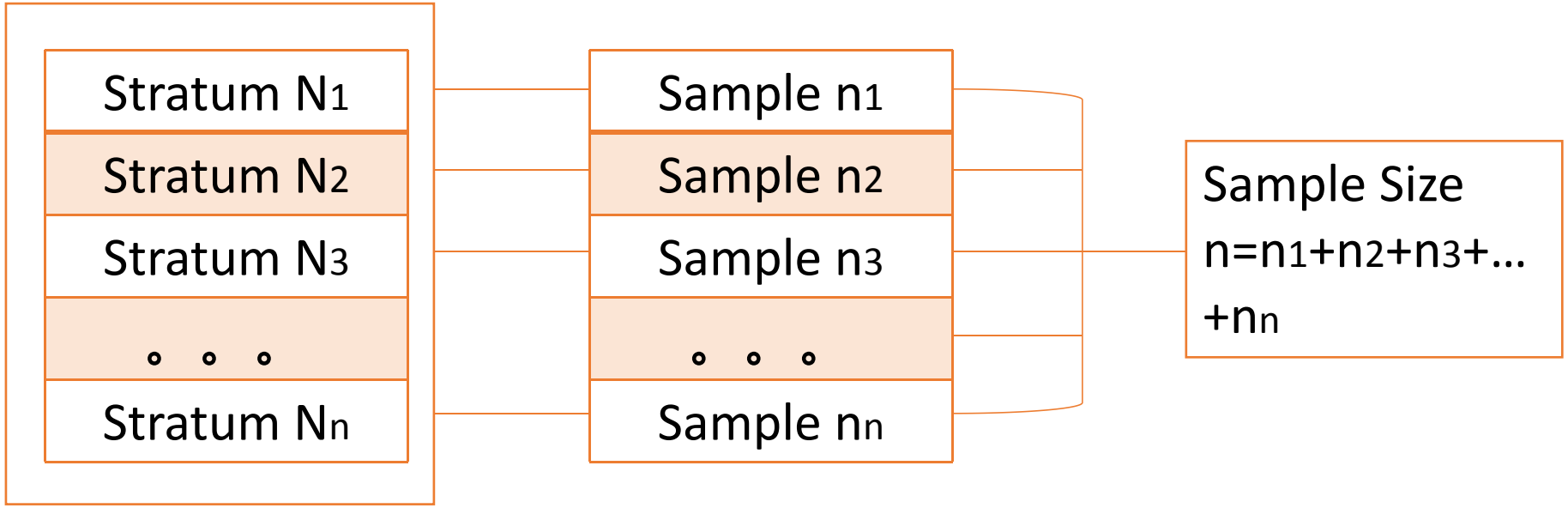
Suppose we want a simple random sample of *size 1000*.

In that case, using a computer *random-number* generator or a table of random numbers.

We then match these random numbers with the number codes of the population members until we have selected a sample of size 1000.

Sampling

- **Stratified Random Sampling**: the population divided into *subpopulations* based on one or more classification criteria. Simple random samples are then drawn from each stratum in sizes proportional to the *relative size of each stratum* in the population. There samples are then pooled to from *a stratified random sample.*



Population
$N=N_1+N_2+..+N_n$

Relative size of each stratum

A Stratified random sample

DATA MINING BIAS

“预言帝”章鱼•保罗

2010年南非世界杯

SURVIVORSHIP BIAS

- In general, hedge funds are **not required to publicly disclose** performance.
- Hedge funds with **poor track records** clearly may not wish to make their records public, creating a problem of self-selection bias in hedge fund databases.
- Many hedge fund databases drop funds that go out of business.

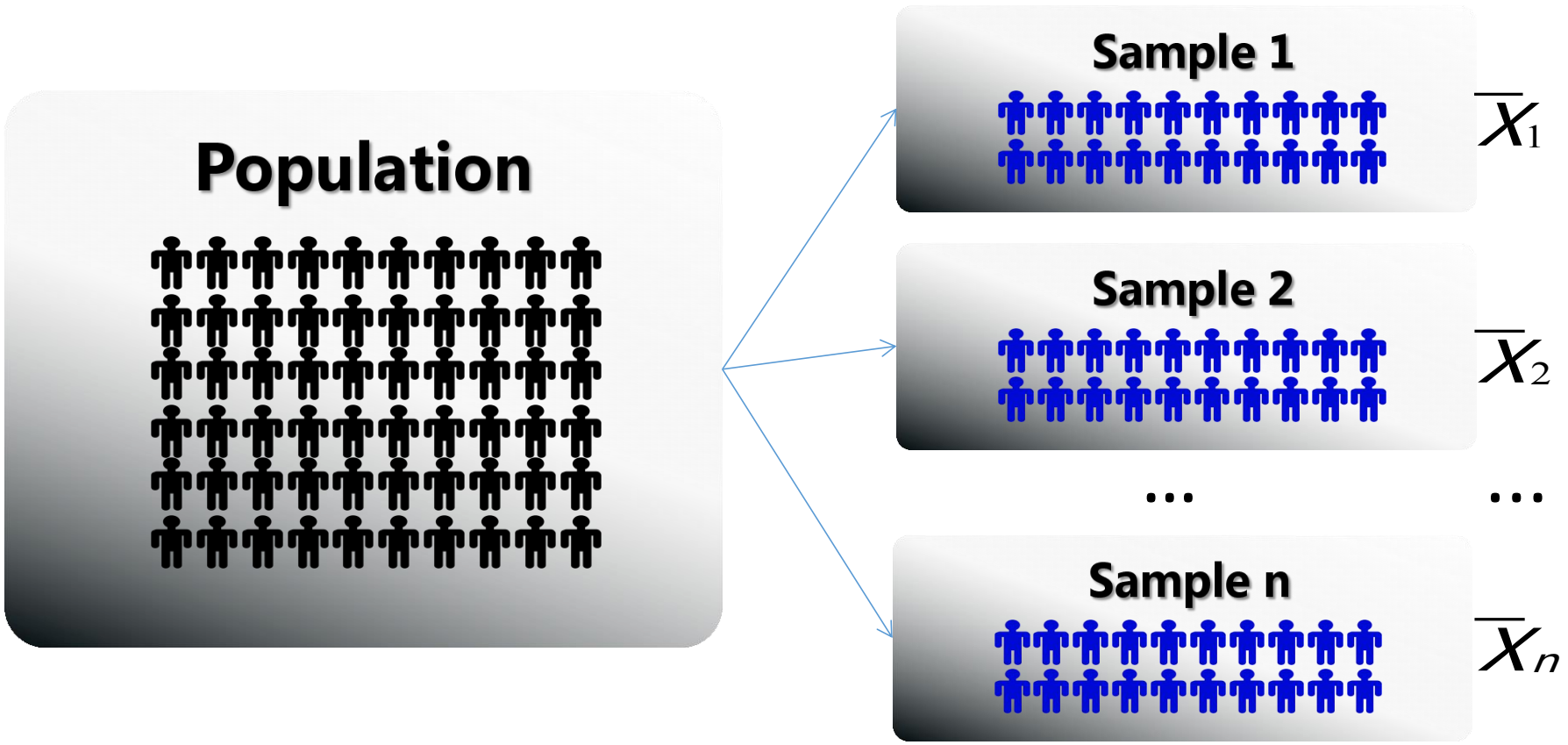
Sampling Distribution

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

Sampling Error

Sampling error is the **difference** between the **sample mean** and the **population mean**

Sampling



总体参数(μ)是常数

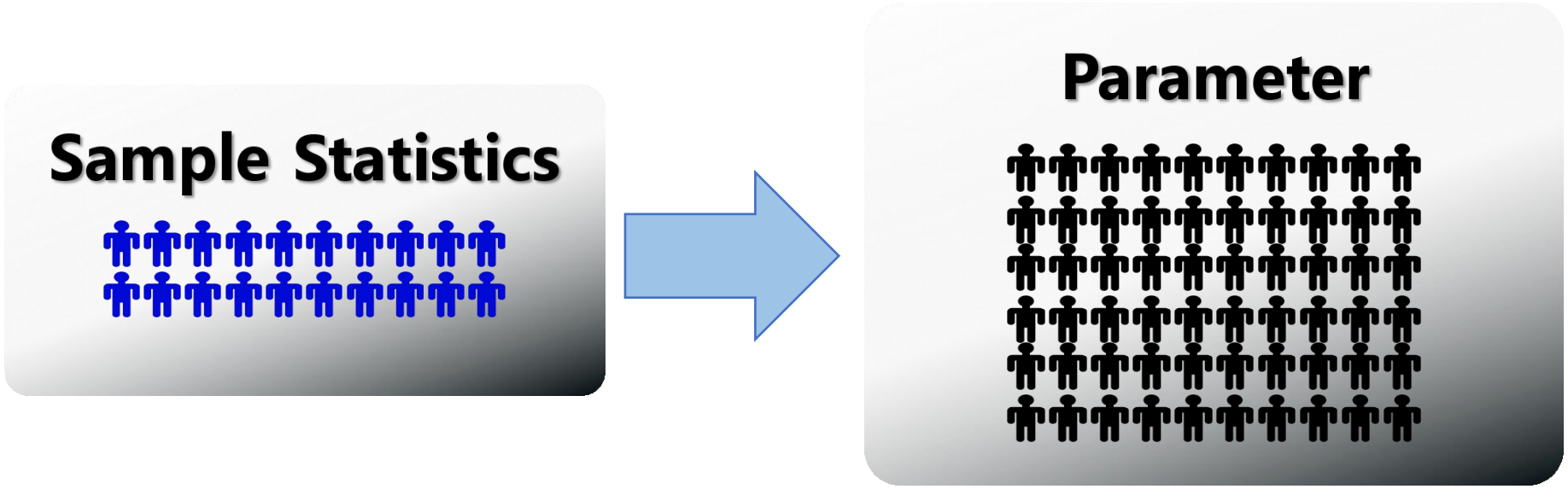
样本统计量是随机变量

Point Estimate

Tasks:

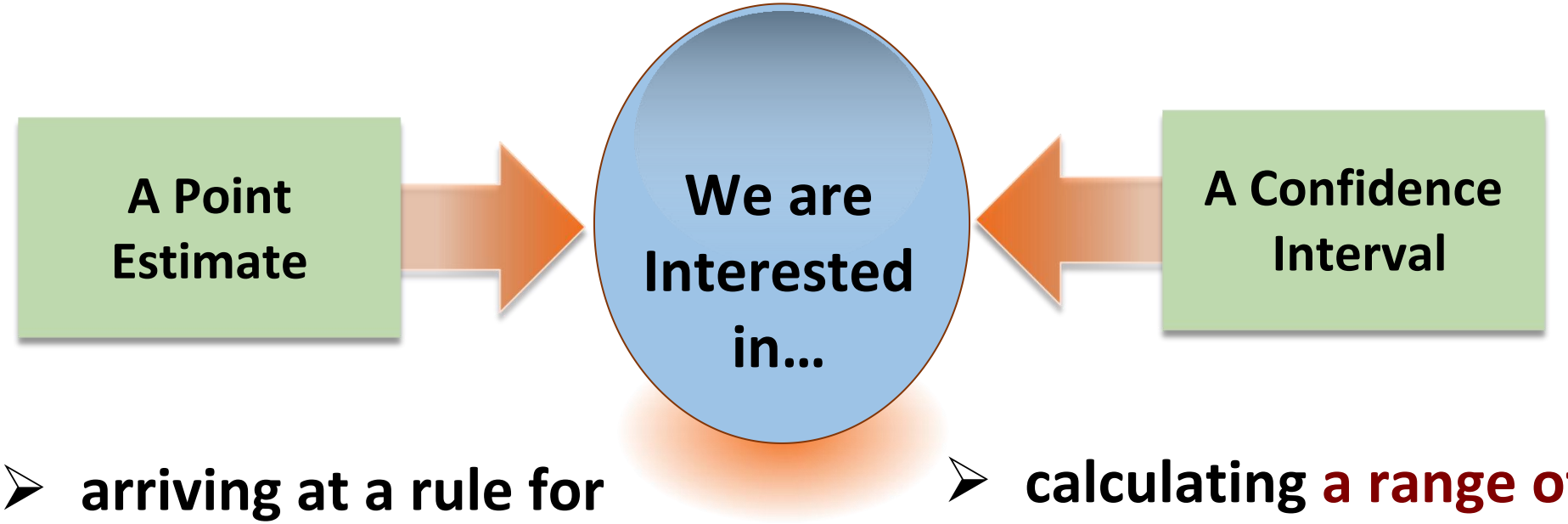
- Describe and interpret the point estimate and confidence interval
- Explain unbiasedness, efficiency and consistency

Estimation



What is this parameter's (for example, the population mean's) value?

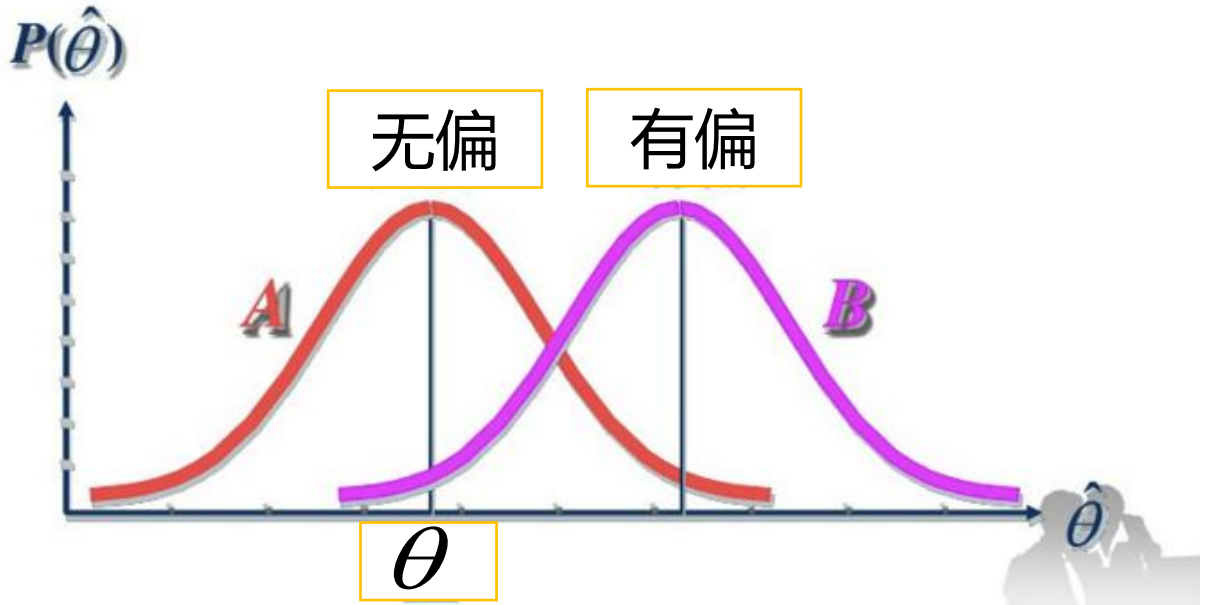
Estimation



- arriving at a rule for best calculating a single number to estimate the unknown population parameter
- calculating a range of values that brackets the unknown population parameter with some specified level of probability.

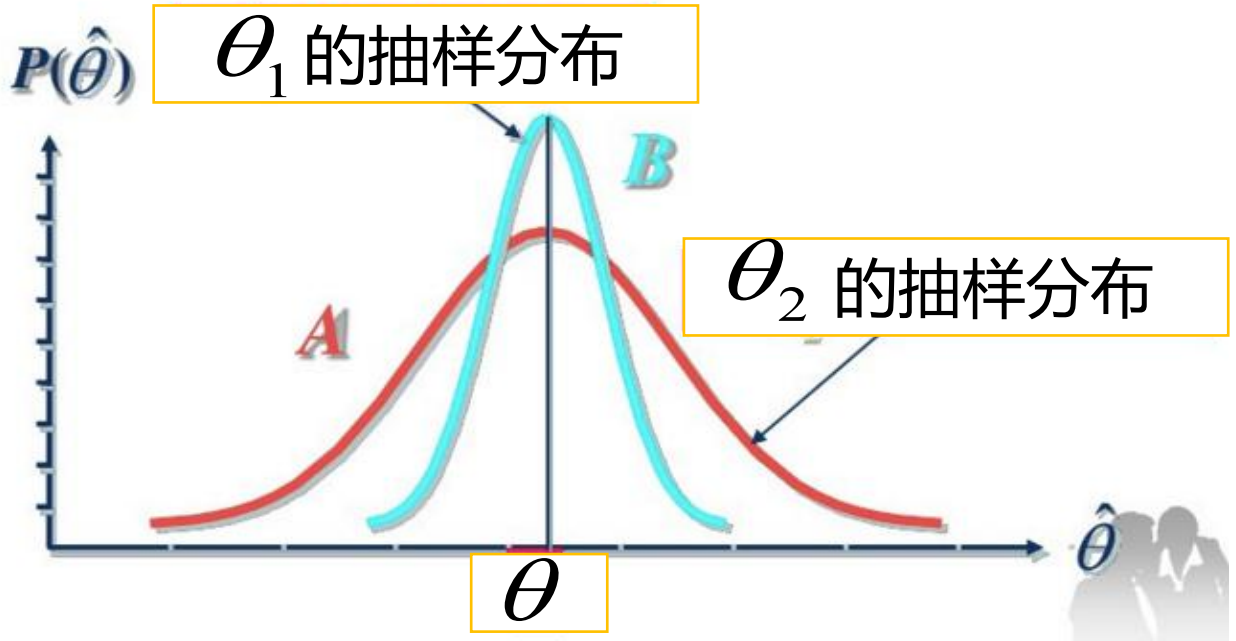
Point Estimate

- **Unbiasedness：**
  - ✓ *An unbiased estimator is one whose expected value (the mean of its sampling distribution) equals the parameter E(A')=A. it is intended to estimate.*
  - ✓ *设A'=g(X1,X2,…Xn)是未知参数A的一个点估计量，若A'满足E（A')=A, 则称A'为A的无偏估计量。*



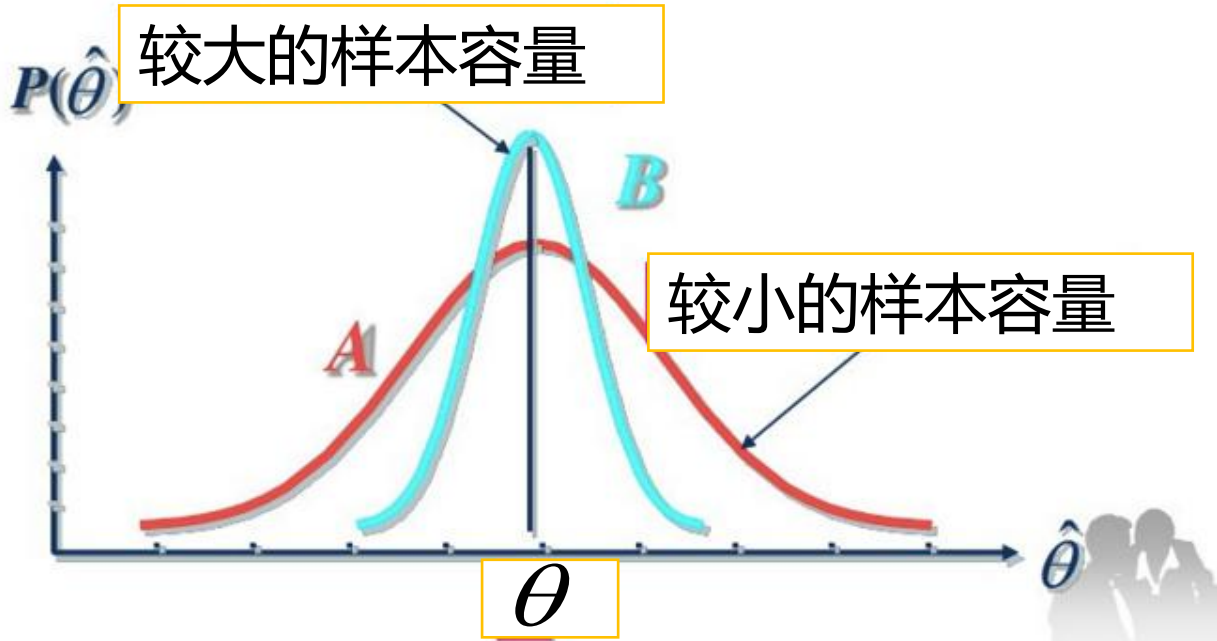
Point Estimate

- **Efficiency：**
  - ✓ *An unbiased estimator is efficient if no other unbiased estimator of the same parameter has a sampling distribution with smaller variance.*
  - ✓ *对同一总体参数的两个无偏估计量，有更小的标准差的估计量更有效。*



Point Estimate

- **Consistency：**
  - ✓ *A consistent estimator is one for which the probability of estimates close to the value of the population parameter increases as sample size increases.*
  - ✓ *随着样本容量的增大，估计量的值越来越接近被估计的总体参数*



Confidence Interval

Tasks:

- Construct a confidence interval
- Explain the Central Limit Theorem

Confidence Interval

Definition:

A confidence interval is a range for which one can assert with a given probability $1 - \alpha$, called the **degree of confidence**, that it will contain the parameter it is intended to estimate. This interval is often referred to as the $100(1 - \alpha)\%$ confidence interval for the parameter.

Construction of Confidence Intervals

A 100(1 – α)% confidence interval for a parameter has the following structure:

Point estimate ± Reliability factor × Standard error

- **Point estimate** = a point estimate of the parameter (a value of a sample statistic)
- **Reliability factor** = a number based on the assumed distribution of the point estimate and the degree of confidence (1 – α) for the confidence interval
- **Standard error** = the standard error of the sample statistic providing the point estimate

Central Limit Theorem

Sample Size (n) is Large

The distribution of the sample mean will be approximately normal.

Approximately Normal Distribution

The mean of the distribution of will be equal to the mean of the population from which the samples are drawn. and Presence

$$E(\overline{X}) = \mu$$

The variance of the distribution of will be equal to the variance of the population divided by the sample size

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Standard Error of the Sample Mean

For sample mean $\overline{X}$ calculated from a sample generated by a population with standard deviation σ, the standard error of the sample mean is given by one of two expressions:

When we know σ

Population Standard deviation

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

When we do not know σ

Need to use the sample standard deviation, s, to estimate it.

$$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

Practice Question

Suppose an investment analyst takes a random sample of US equity mutual funds and calculate the average Sharpe ratio. The sample size is 100, and the average Sharp ratio is 0.45. The sample has a standard deviation of 0.3. Calculate and interpret the 90 percent confidence interval for the population mean of all US equity mutual funds by using a reliability factor based on the standard normal distribution.

TIPS: The reliability factor for a 90 percent confidence interval is z0.05 = 1.65.

Answer: The confidence interval will be

$$\overline{X} \pm z_{0.05} \frac{s}{\sqrt{n}} = 0.45 \pm 1.65 \frac{0.3}{\sqrt{100}} = 0.45 \pm 1.65(0.03) = 0.45 \pm 0.0495$$

The confidence interval spans 0.4005 to 0.4995. The analyst can say with 90 percent confidence that the interval includes the population mean.

Hypothesis test

Tasks:

- Describe the process of hypothesis test
- Construct the Null hypothesis and Alternative hypothesis
- Describe the Type I and Type II error

Hypothesis Test

“中国人的平均身高（μ）是160cm”

“不！你说的不对！！！”

“那就证明给我看吧”

“我做了抽样检验，样本量（n）为1000，抽样显示，样本均值（$\bar{x}$）为175cm，标准误（$s_{\bar{x}}$）为5cm”

“根据中心极限定理，样本均值服从正态分布。如果你说的是对的（μ=160），那我有95%的把握样本均值（$\bar{x}$）应该在150~170cm之间。但现在是175cm，不在这个区间范围内，所以你说中国人平均身高为160cm，应该是不对的。”

Estimation vs. Hypothesis Test

➢ Suppose the sample mean is 50 and a 95% confidence interval for the population mean is 50±10 (the confidence interval runs from 40 to 60).

Estimation
"What is this parameter's value?"

PK

Hypothesis Test
"Is the value of the parameter 50?"

If this confidence interval has been ***properly constructed***, there is a 95% probability that the interval from 40 to 60 ***contains the population mean's value***.

The assertion "***the population mean is 50***" is a hypothesis. we seek to ***determine whether*** the evidence from a sample supports or does not support that hypothesis.

Hypothesis Test

导师转身

导师没有转身

姚贝娜，她是不是好歌手？

State the hypotheses

导师：我该如何做抉择？

State the decision rule

Make the statistical decision

Hypothesis Test

Null Hypothesis

- Is the hypothesis to be tested.
- Is a proposition that is considered true unless the sample we use to conduct the hypothesis test gives convincing evidence that the null hypothesis is false.

Alternative Hypothesis

- The alternative hypothesis accepted when the null hypothesis is rejected.
- Alternative has the **'hoped for'** condition.

When such evidence is present, we are led to the alternative hypothesis.

Hypothesis Test

Formulations of Hypotheses

1. **$H_0: \theta = \theta_0$ vs. $H_a: \theta \neq \theta_0$**
   (a '*not equal to*' alternative hypothesis)
2. **$H_0: \theta \leq \theta_0$ vs. $H_a: \theta > \theta_0$**
   (a '*greater than*' alternative hypothesis)
3. **$H_0: \theta \geq \theta_0$ vs. $H_a: \theta < \theta_0$**
   (a '*less than*' alternative hypothesis)

Example: Susan Bellows is comparing the return on equity for two industries. She is convinced that the return on equity for the discount retail industry(DR) is greater than that of the luxury retail(LR) industry. What are the hypotheses for a test of her comparison of return on equity?

$$H_o: \mu_{DR} \leq \mu_{LR} \ \textit{versus}\ H_a: \mu_{DR} > \mu_{LR}$$

Hypothesis Test

Two actions are possible for the decision making

01 We reject the null hypothesis

02 We do not reject the null hypothesis

There are four possible outcomes when we test a null hypothesis

We reject a false null hypothesis. This is a power of a test.

We do not reject a false null hypothesis. This is called a *Type II error*.

We reject a true null hypothesis. This is called a *Type I error*.

We do not reject a true null hypothesis. This is called *correct decision*

Hypothesis Test

Decision	True Situation	
	H_0 True 他/她是好歌手	H_0 False 他/她不是好歌手
Do not reject H_0 导师转身	Correct Decision	Type II Error
Reject H_0(Accept Ha) 导师不转身	Type I Error	Power of Test

Example: which of the following statements about hypothesis testing is most accurate? A Type I error is the probability of:

A. Failing to reject a false hypothesis

B. Rejecting a true null hypothesis

C. Rejecting a true alternative hypothesis

***Answer B**; The Type I error is the error of rejecting the null hypothesis when, in fact, the null is true.*

You're a Champion!

Thanks for staying with us. You have finished this chapter.