**Contents**

[Chapter 1 Quantitative Methods 数量分析 1](#_Toc61480784)

[1.1 Reading 10 Sampling and Estimation 抽样估计 1](#_Toc61480785)

[1.1.1 Sampling 抽样 2](#_Toc61480786)

[1.1.2 Types of Data 数据分类 2](#_Toc61480787)

[1.1.3 Central Limit Theory 中心极限定理 ☆☆ 2](#_Toc61480788)

[1.1.4 Properties of Estimator 估计量的性质 3](#_Toc61480789)

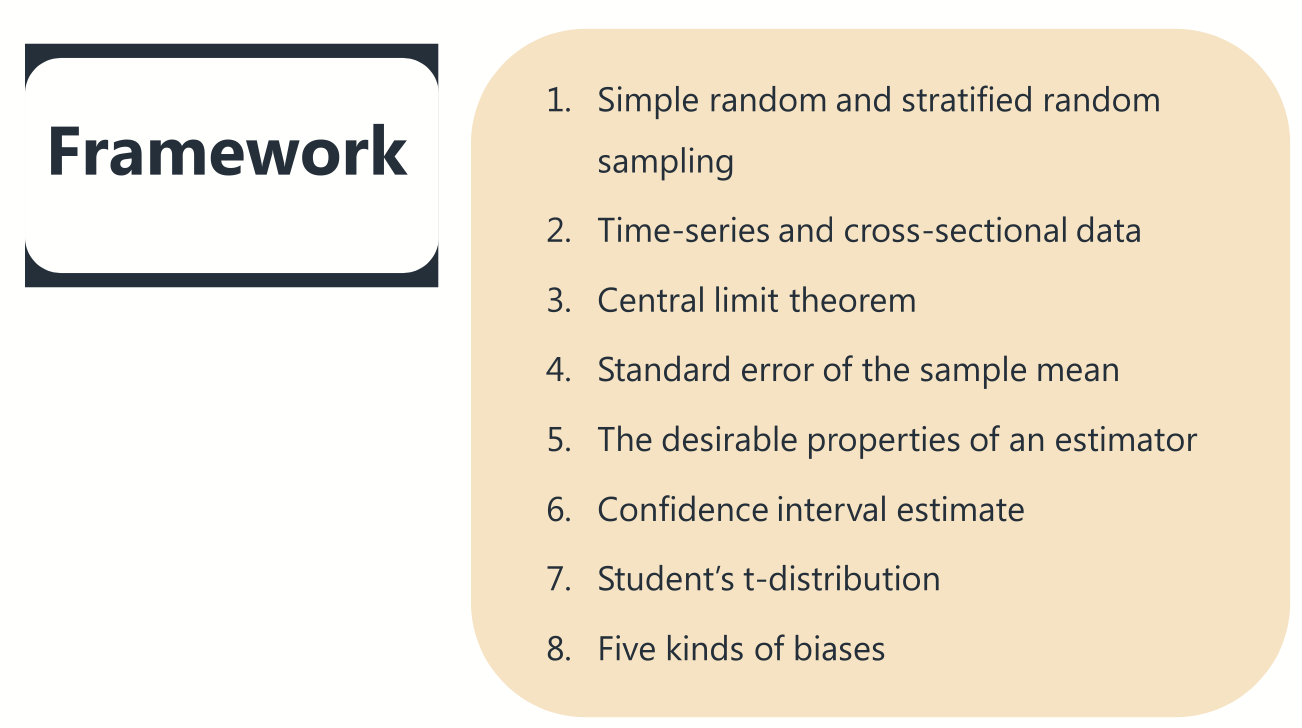
[1.1.5 Estimation 估计 3](#_Toc61480790)

[1.1.6 T-Distribution t分布 3](#_Toc61480791)

[1.1.7 Types of Bias 各种偏差 4](#_Toc61480792)

# Quantitative Methods 数量分析

## Reading 10 Sampling and Estimation 抽样估计



### Sampling 抽样

* 抽样方法
  + Simple random sampling 简单随机抽样法
  + Stratified random sampling分层随机抽样法
* Sampling error 抽样误差
  + Sampling error of the mean=sample mean-population mean 抽样误差=样本统计值-总体值

n sampling error

* Sample statistic 样本统计量：描述样本的特征量random variable 随机变量：取值不确定的量probability distribution 概率分布

### Types of Data 数据分类

* Time-series data 时间序列数据

同一公司不同时间的一系列数据

* Cross-sectional data 横截图数据

同一时间不同公司的一系列数据

### Central Limit Theory 中心极限定理 ☆☆

The sampling distribution of the sample mean approaches N(, /n) if the sample size is sufficient large(n).

* Central Limit Theory
  + 条件

1. n
2. 总体均值、方差已知
   * 结论
3. 服从正态分布
4. , Var()=

* Standard error of the sample mean
  + Known population variance

sample standard deviation 样本标准差 sample mean standard error 标准误

* + Unknown population variance

### Properties of Estimator 估计量的性质

* The desirable properties of an estimator 估计量比较好的性质
  + Unbiasedness 无偏性
  + Efficiency 有效性 Var最小
  + Consistency 一致性

### Estimation 估计

* 抽样推断的三种方法
  + Point estimate 点估计
  + Confidence interval estimate CI 置信区间估计

1. 计算
2. Width:
   * Interval Estimation 区间估计
3. Level of Significance (alpha) 显著性水平
4. Degree of Confidence (1-alpha) 置信度

### T-Distribution t分布

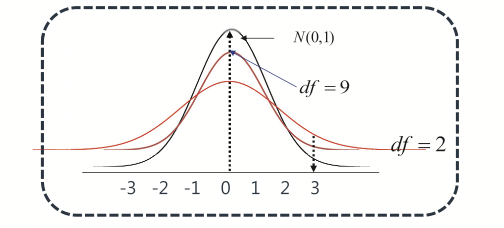
* Student’s t-distribution
  + Symmetricalskewness=0
  + Degree of freedom(df): n-1
  + Less peaked than a normal distribution(“fatter tails”)

T分布和正态分布相比，低峰肥尾(kurtosis<3)，方差更大(

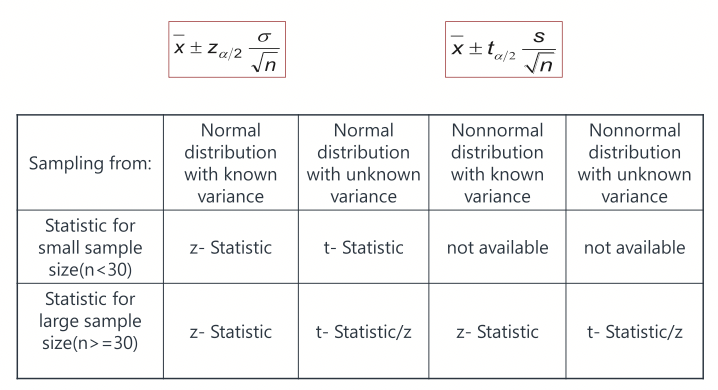
* + Student’s t-distribution converges to the standard normal distribution as degrees of freedom goes to infinity.

df

* + 相同，CIt更宽



* t分布与z分布
  + 方差已知用z
  + 方差未知用t
  + 非正态总体小样本不可估计
  + 样本容量足够大，任何情况均可用z



### Types of Bias 各种偏差

* 五种偏差
  + Data-mining bias 把偶然当成必然
  + Sample selection bias 样本选择性偏差
    - Survivorship bias 生存性偏差
  + Look-ahead bias 前视性偏差
  + Time-period bias 时间段偏差