

第1章 导论

内容提要

1.1 统计学的概念

1.2 统计学的分类

1.3 统计学的应用领域

1.4 统计数据类型

1.5 统计学中的几个基本概念

1.1 统计学的概念

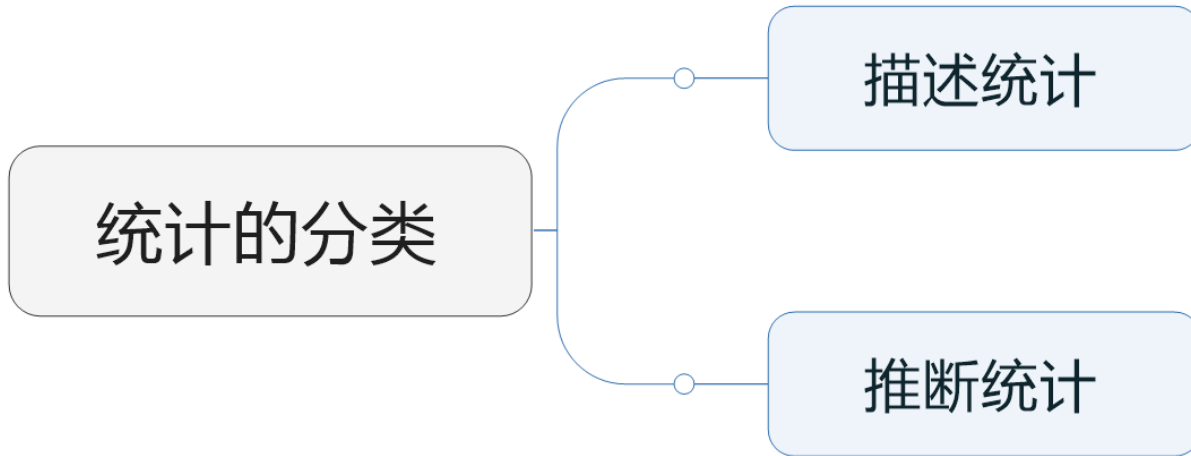
- **统计学**是收集、分析、表述和解释**数据**的科学。(不列颠百科全书)
- **统计学**是一门收集、分析、解释和提供**数据**的科学。(韦伯斯特国际辞典第3版)
- **统计学**指的是一组方法，用来设计实验、获得**数据**，然后在这些数据的基础上组织、概括、演示、分析、解释和得出**结论**。(Mario F.Triola, 《初级统计学》)

1.1 统计学的概念

• **统计学**：收集、处理、分析、解释**数据**并从数据中得出**结论**的科学。（课本的定义）

1. 收集数据：取得数据
2. 处理数据：整理与图表展示
3. 分析数据：利用统计方法分析数据
4. 解释数据：结果的说明
5. 得到结论：从数据分析中得出客观结论

1.2 统计学的分类



1.2 统计学的分类

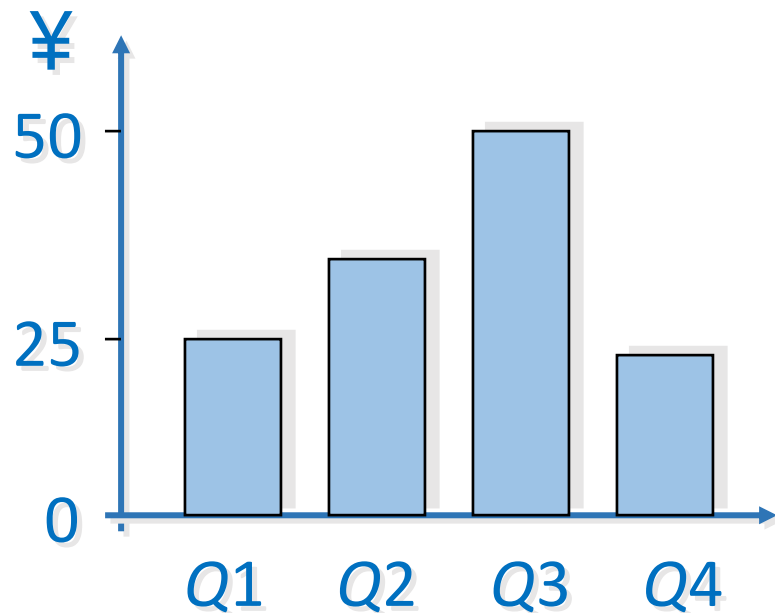
- **描述统计(descriptive statistics)**: 研究数据收集、处理、汇总、图表描述、概括与分析等统计方法。

- 内容

- 搜集数据
- 整理数据
- 展示数据
- 描述性分析

- 目的

- 描述数据特征
- 找出数据的基本规律



$$\bar{x}=30 \quad s^2=105$$

1.2 统计学的分类

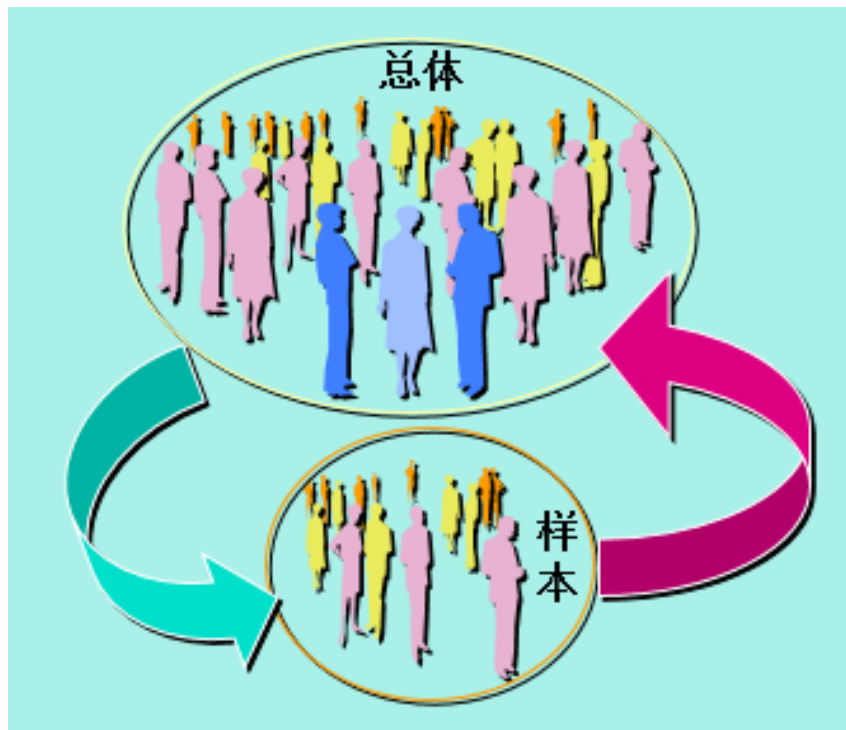
- **推断统计(inferential statistics)**: 研究如何利用**样本**数据来**推断**总体特征的统计方法。

- 内容

- 参数估计
- 假设检验

- 目的

- 对总体特征作出推断



1.3 统计学的应用领域

- actuarial work (精算)
- animal science (动物学)
- archaeology (考古学)
- crystallography (晶体学)
- dentistry (牙医学)
- election forecasting and projection (选举预测和策划)
- engineering (工程)
- epidemiology (流行病学)
- fisheries research (水产渔业研究)
- genetics (遗传学)
- **historical research (历史研究) —— 量化历史**
- human genetics (人类遗传学)
- agriculture (农业)
- anthropology (人类学)
- auditing (审计学)
- demography (人口统计学)
- **econometrics (计量经济学)**
- ecology (生态学)
- education (教育学)
- finance (金融)
- gambling (赌博)
- geography (地理学)
- geology (地质学)
- hydrology (水文学)

1.3 统计学的应用领域

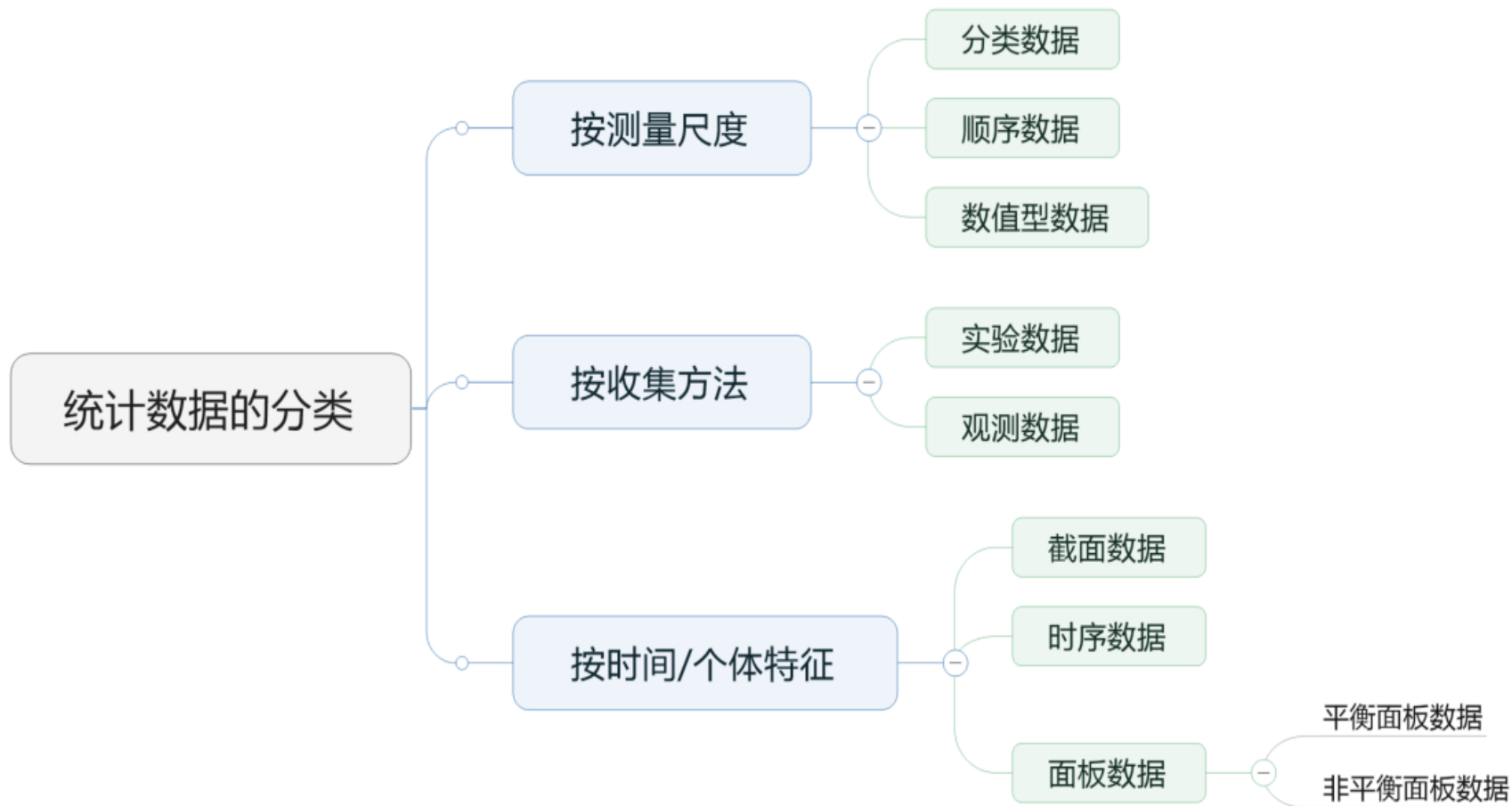
- industry (工业)
- literature (文学)——[“红学”研究中的统计应用](#)
- manpower planning (劳动力计划)
- management science (管理科学)
- medical diagnosis (医学诊断): [血氧饱和度](#) [新冠诊疗方案 \(第8版\)](#)
- meteorology (气象学)
- nuclear material safeguards (核材料安全管理)
- pharmaceuticals (制药学): [疫苗比较](#), [有效性计算](#), [临床三期](#)
- physics (物理学)
- psychology (心理学)
- quality control (质量控制)
- sociology (社会学)
- weather modification (气象改善)
- linguistics (语言学)
- marketing (市场营销学)
- military science (军事科学)
- ophthalmology (眼科学)
- political science (政治学)
- psychophysics (心理物理学)
- religious studies (宗教研究)
- survey sampling (调查抽样)
- taxonomy (分类学)

1.3 统计学的应用领域

• 经管领域的统计应用

- 企业发展战略
- 产品质量管理
- 市场研究
- 财务分析
- 经济预测
- 人力资源管理

1.4 统计数据的类型



1.4 统计数据的数据类型

➤ 按测量尺度

- **分类数据(categorical data)**

- 只能归于某一类别的非数字型数据
- 对事物进行分类的结果，数据表现为类别，用文字来表述
- 例如，人口按性别分为男、女两类，如何转换为数据？

- **顺序数据(rank data)**

- 只能归于某一有序类别的非数字型数据
- 对事物类别顺序的测度，数据表现为类别，用文字来表述
- 例1，受教育程度：未上学、小学、初中、高中，大学等，如何转换为数据？
- 例2：满意度通常分为：非常不满意，不满意，无所谓，满意，很满意，如何转换为数据？测量是否精确？

- **数值型数据(metric data)**

- 按数字尺度测量的观察值
- 结果表现为具体的**数值**和**单位**（**测量尺度的体现**）
- 例如：身高为175cm、168cm、183cm

1.4 统计数据的数据类型

测量尺度的类型、属性及变量类型对应表

		测量尺度属性			变量类型
		顺序属性	距离属性	比率属性	
测量 尺度 类型	比率尺度	√	√	√	数值型变量
	区间尺度	√	√	×	
	序列尺度	√	×	×	顺序型变量
	名义尺度	×	×	×	分类型变量

注：“√”表示测量尺度具备该属性，“×”表示测量尺度不具备该属性。

1.4 统计的数据类型

➤ 按收集方法

- **观测数据(observational data)**

- 通过调查或观测而收集到的数据
- 在**没有对事物人为控制**的条件下而得到的
- 有关**社会经济现象**的统计数据几乎都是观测数据
- 例子: [统计数据是如何产生的: 国内生产总值 \(GDP\)](#)

- **实验数据(experimental data)**

- 在实验中**控制实验对象**而收集到的数据
- 比如, 对一种新药疗效的实验, 对一种新的农作物品种的实验等
- **自然科学领域**的数据大多数都为实验数据

1.4 统计数据的类型

➤按时间/个体特征

• 截面数据(cross-sectional data)

- 在**相同的时间点**上收集的 (**多个个体**) 数据
- 描述现象在某一时刻的变化情况
- 例子: 2011年中国各省GDP数据

• 时间序列数据(time series data)

- 在**不同时间**上收集到的 (**单一个体**) 数据
- 描述现象随时间变化的情况
- 例子: 2011年至2015年北京市GDP数据

• 面板数据 (panel data)

- 在连续**不同的时间点**上, 收集的**多个个体**的数据
- 根据是否存在缺漏, 可区分为:
 - 平衡面板(balance panel)数据
 - 非平衡面板(unbalance panel)数据
- 例子: 2011-2015年中国各省的GDP数据

3-9 地区生产总值和指数

本表绝对数按当年价格计算, 指数按不变价格计算。

地区	地区生产总值(亿元)				
	2011	2012	2013	2014	2015
北京	16251.93	17879.40	19800.81	21330.83	23014.59
天津	11307.28	12893.88	14442.01	15726.93	16538.19
河北	24515.76	26575.01	28442.95	29421.15	29806.11
山西	11237.55	12112.83	12665.25	12761.49	12766.49
内蒙古	14359.88	15880.58	16916.50	17770.19	17831.51
辽宁	22226.70	24846.43	27213.22	28626.58	28669.02
吉林	10568.83	11939.24	13046.40	13803.14	14063.13
黑龙江	12582.00	13691.58	14454.91	15039.38	15083.67
上海	19195.69	20181.72	21818.15	23567.70	25123.45
江苏	49110.27	54058.22	59753.37	65088.32	70116.38
浙江	32318.85	34665.33	37756.58	40173.03	42886.49
安徽	15300.65	17212.05	19229.34	20848.75	22005.63
福建	17560.18	19701.78	21868.49	24055.76	25979.82
江西	11702.82	12948.88	14410.19	15714.63	16723.78
山东	45361.85	50013.24	55230.32	59426.59	63002.33
河南	26931.03	29599.31	32191.30	34938.24	37002.16
湖北	19632.26	22250.45	24791.83	27379.22	29550.19
湖南	19669.56	22154.23	24621.67	27037.32	28902.21
广东	53210.28	57067.92	62474.79	67809.85	72812.55
广西	11720.87	13035.10	14449.90	15672.89	16803.12
海南	2522.66	2855.54	3177.56	3500.72	3702.76
重庆	10011.37	11409.60	12783.26	14262.60	15717.27
四川	21026.68	23872.80	26392.07	28536.66	30053.10
贵州	5701.84	6852.20	8086.86	9266.39	10502.56
云南	8893.12	10309.47	11832.31	12814.59	13619.17
西藏	605.83	701.03	815.67	920.83	1026.39
陕西	12512.30	14453.68	16205.45	17689.94	18021.86
甘肃	5020.37	5650.20	6330.69	6836.82	6790.32
青海	1670.44	1893.54	2122.06	2303.32	2417.05
宁夏	2102.21	2341.29	2577.57	2752.10	2911.77
新疆	6610.05	7505.31	8443.84	9273.46	9324.80

1.5 统计中的几个基本概念

- **总体(population)**

- 所研究的全部个体的集合，其中的每一个个体称为元素
- 分类：有限总体和无限总体
 - **有限总体**的范围能够明确确定，且元素的数目是有限的；**非独立抽样**
 - **无限总体**所包括的元素是无限的，不可数的；**独立抽样**

- **样本 (sample)**

- 从总体中抽取的一部分元素的集合；
- 构成样本的元素的数目称为样本容量或样本量 (sample size)

1.5 统计中的几个基本概念

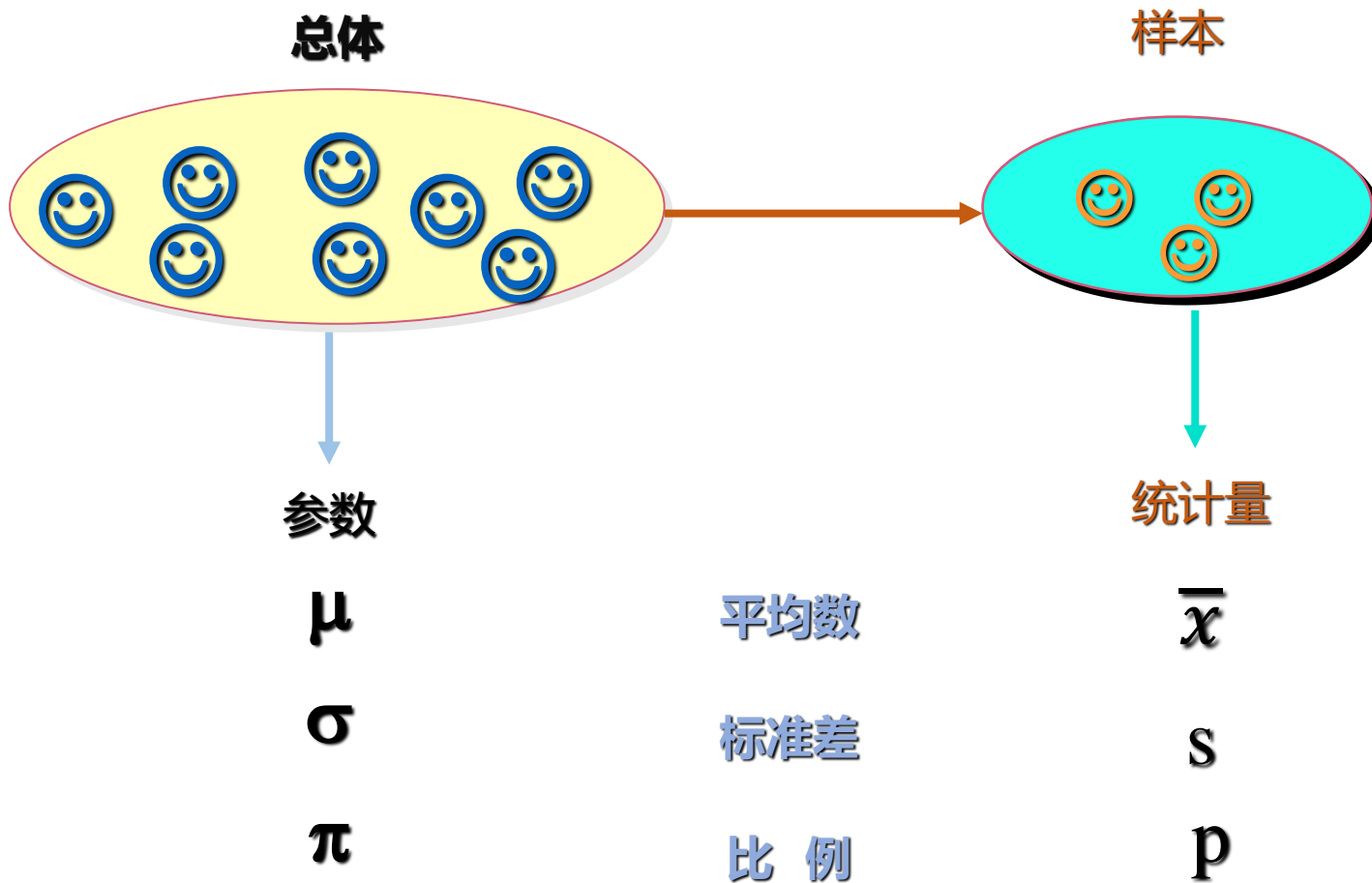
- **参数(parameter)**

- 描述**总体**特征的概括性数字度量，是研究者想要了解的总体的某种特征值
- 所关心的**参数**主要有**总体均值(μ)**、**标准差(σ)**、**总体比例(π)**等
- 总体参数通常用希腊字母表示

- **统计量(statistic)**

- 用来描述**样本**特征的概括性数字度量，它是根据样本数据计算出来的一些量，是样本的函数
- 所关心的样本**统计量**有**样本均值(\bar{x})**、**样本标准差(s)**、**样本比例(p)**等
- 样本统计量通常用小写英文字母来表示

1.5 统计中的几个基本概念



1.5 统计中的几个基本概念

- **变量的定义**：说明现象某种特征的概念。
- **区分：变量与数据**
 - 数据——经验变量（实证变量/统计指标）的某些取值；
- **变量的分类**：
 - 按**测量尺度**：
 - 分类变量(categorical variable)：说明事物**类别**的名称
 - 顺序变量(rank variable)：说明事物**有序类别**的名称
 - 数值型变量(metric variable)：说明事物**数字特征**的名称
 - 离散变量：取有限个值
 - 连续变量：可以取无穷多个值

1.5 统计中的几个基本概念

- 变量的分类（续）：

- 按**抽象程度**：

- 理论变量——基于数学方法，或理论构造出的变量；

- 例如：z统计量、t统计量等；

- 经验变量（实证变量/统计指标）——描述可观测事物的变量；

- *对应关系：某个理论变量可以对应多个经验变量（统计指标）*

- *例1：家庭收入—家庭可支配收入，或家庭人均收入等；*

- *例2：受教育程度—受教育年限，或受教育层次*

- 按**是否随机**：

- 随机变量—样本

- 非随机变量—总体

小结：变量分类



注：在统计学原理后续的课程中，如未作特殊说明，变量一般是指**经验变量（实证变量，或统计指标）**。

本章小结

