

《统计学》

第一章 导论

统计 是“认识社会的最有力的武器之一”
——列宁

Statistical thinking will one day be as necessary
for efficient citizenship as the ability to read
and write.

—— H.G.Wells (英国作家、历史学家)



- 1.经最终核实，2019年，GDP现价总量为986515亿元，比初步核算数减少了4350亿元；按不变价格计算，比上年增长6.0%，比初步核算数下降0.1个百分点。三次产业和各行业数据见附表。

——国家统计局

- 2.吸烟对健康是有害的。



第一章 导论

第一节 统计及其应用领域

第二节 统计学中的基本概念

第三节 SPSS及Excel基本操作



第一节 统计及其应用领域

- 一、什么是统计学
- 二、统计的应用领域
- 三、统计研究过程



一、什么是统计学

统计的三种含义

统计活动

对各种统计数据进行收集、整理并做出相应的推断、分析的活动，通常可以划分为统计调查、统计整理和统计分析三个阶段

统计数据

通过统计活动获得的，用以表现研究现象特征的各种形式的信息(数值、文字、图像)

统计学

是指导统计活动的理论和方法，是收集、处理、分析、解释数据，并从数据中得出结论的一门通用的方法论科学

统计数据与统计活动是统计结果与过程的关系，统计活动与统计学是统计实践与理论的关系。



一、什么是统计学

从方法论角度看，统计学方法通常可以概括为：

描述统计方法

研究如何取得反映客观现象的数据，并通过图表形式对所收集的数据进行加工处理和显示，进而通过综合、概括与分析，得出反映客观现象的规律性的数量特征。

推断统计方法

研究如何根据样本数据去推断总体的数量特征，它是在对样本数据进行描述的基础上，对统计总体的未知数量特征做出以概率形式表述的推断。



一、什么是统计学

描述统计方法主要包括：

- 数据的收集方法
- 数据的加工处理方法
- 数据的显示方法
- 数据分布特征的概括与分析方法



二、统计的应用领域

统计的应用领域：

- 公共管理领域 使决策科学化
- 经济研究领域 经济研究定量化和数理化
- 产品质量管理 生产过程的监测活动
- 市场调研领域 搜集和分析市场信息
- 人力资源管理 设计合理的绩效考评机制
- 金融保险领域 处理相关数据



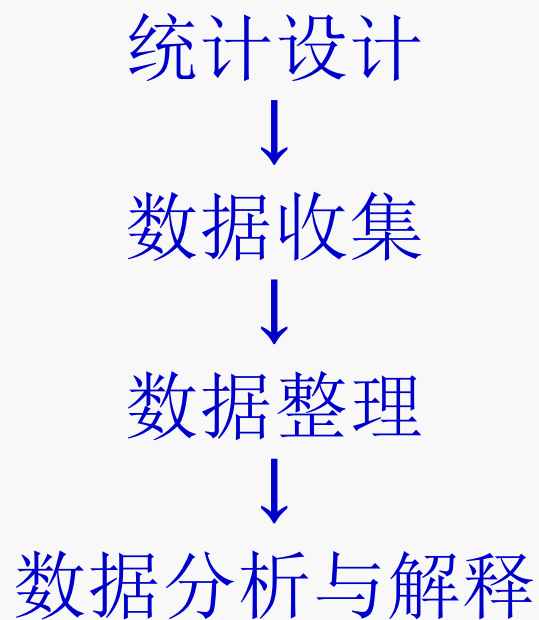
二、统计的应用领域

- 统计学在医学领域、考古领域、心理学领域、气象学领域、物理学领域及生物学领域等也都发挥着重要的作用
- 统计不是万能的，并不能解决人们面临的所有问题：
 - 统计方法帮助人们分析数据得出某种结论，但需要相应的专业知识对统计结论进一步解释说明（吸烟与肺癌）



三、统计研究过程

统计研究过程包括：



三、统计研究过程

统计设计 对整个统计研究过程所涉及的各个环节进行的通盘考虑和事先安排

统计设计的内容

- 确定统计研究的对象
- 明确所要收集数据的种类
- 确定相应的统计指标及其体系
- 确定统计数据收集、整理、推断和分析的基本方法
- 规定研究工作的进度安排和质量要求
- 拟定研究工作的资源配置和组织实施方式等



三、统计研究过程

数据收集

是按照统计设计的要求，通过统计观测或实验的方式、方法，收集所需的原始数据以及其它已经存在的各种相关数据的过程

- 数据收集是否准确、及时、完整，将直接影响到统计分析的质量



三、统计研究过程

数据整理

是对通过统计观测或实验所获得的原始数据，进行必要的系统化处理，使之条理化、综合化，成为能够反映总体特征的统计数据的过程

数据分析与解释

是在数据整理结果的基础上，围绕统计设计所确定的研究任务，运用各种统计方法对数据进行分析，得出某些有用的定量结论的过程



第二节 统计学中的基本概念

- 一、总体与样本
- 二、参数与统计量
- 三、标志与统计指标
- 四、变量及其测量尺度



一、总体与样本

总体

指包含所研究对象的全部个体（数据）的集合

总体单位 (个体)

指组成总体的每一个元素

- 总体由总体单位构成，要认识总体必须从总体单位开始，总体是统计认识的对象
- 总体或总体单位的区分不是固定的，在一定条件下可以相互转化



一、总体与样本

总体

有限总体

总体的范围能够明确界定，所包含个体的数目是有限可数的

无限总体

所包含个体的数目是无限的、不可数的

- 把总体分为有限总体和无限总体，主要是为了判别在抽样中每次抽取是否独立。抽取个体的过程是否独立，会影响到抽样推断的结果(详见第五章)



一、总体与样本

样本

是从统计总体中抽取出来作为代表这一总体的、由部分个体组成的集合

样本量 构成样本的个体数目

- 统计学的研究目的是把握研究总体的特征。当总体所包含的个体单位数目很多时，可以抽取总体中的部分个体组成样本，根据样本提供的信息推断总体的特征。



二、参数与统计量

参数

用来描述总体特征的概括性数字度量，通常为未知常数。比如：总体平均数 μ 、总体标准差 σ 、总体比例 Π 、相关系数 ρ 等

统计量

用来描述样本特征的概括性数字度量，根据样本数据计算得到。比如，样本平均数 $\bar{x}$ 、样本标准差 s 、样本比例 p 等

- 参数通常是未知常数，统计量是样本的函数
- 抽样的目的是根据样本统计量去估计总体参数



三、标志与统计指标

统计标志

指每个个体所共同具有的属性和特征，是说明个体的属性或特征的具体名称；标志的具体表现称为**标志值**

- 统计研究从登记标志开始，并通过对标志的综合来反映总体的数量特征，因此标志是统计研究的起点

品质标志

表明事物“**质**”的特性的标志，用文字说明
如：企业性质，所属行业，性别等

数量标志

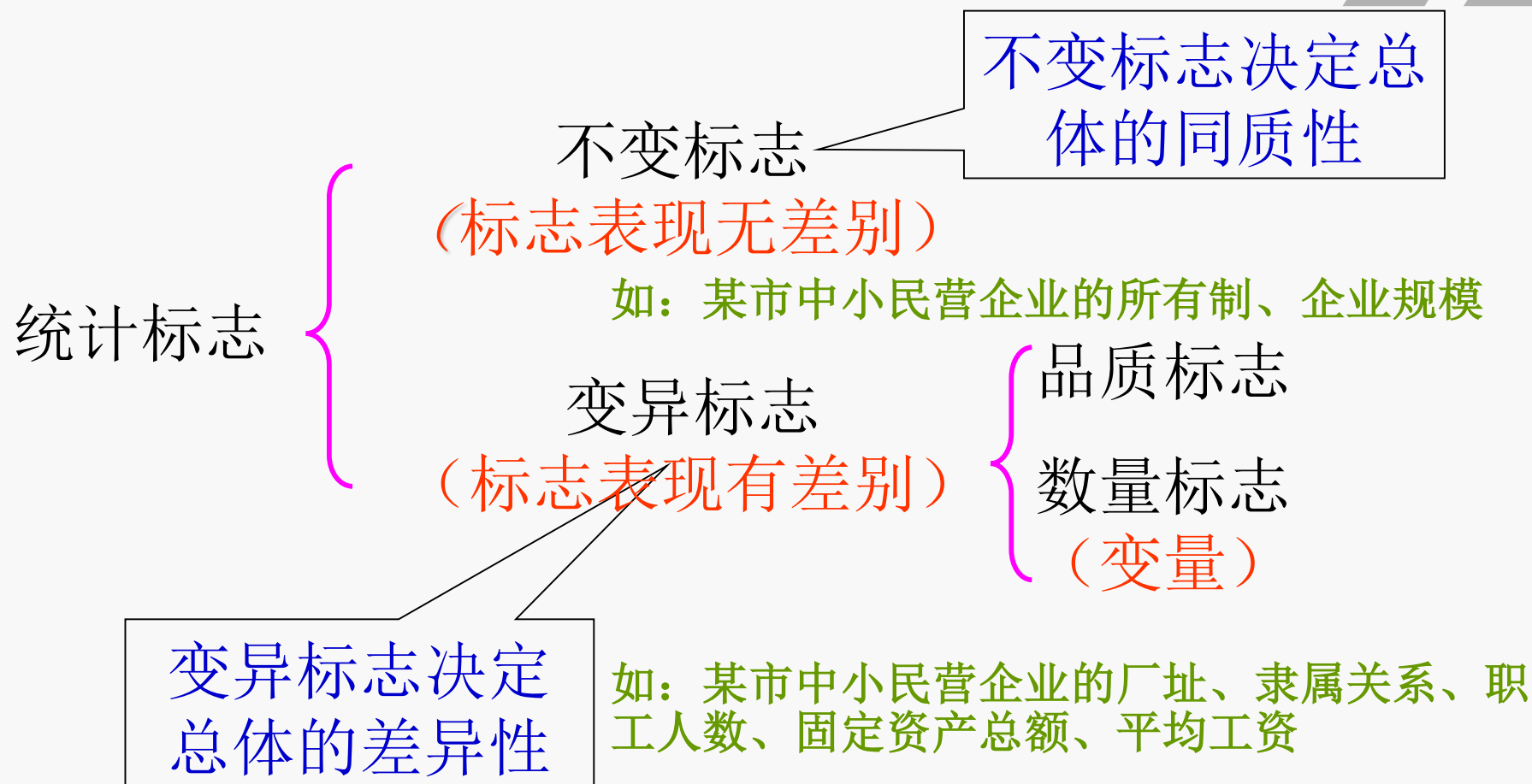
表明事物“**量**”的特性的标志，用数字表示
如：职工人数，产值，身高，体重等



其中，可变的数量标志又被叫作**变量**



三、标志与统计指标



是非标志（交替标志、0-1标志）：用“是”或“否”、“有”或“无”来表示的标志



三、标志与统计指标

统计指标

反映社会经济现象总体数量特征的概念及其具体数值，由各个个体的标志值汇总综合而成。

构成要素： 指标名称 + 指标数值

时间
限制

空间
限制

计算方法
指标名称

具体
数值

计量
单位

2011年末

大连市

户籍人口

588.5

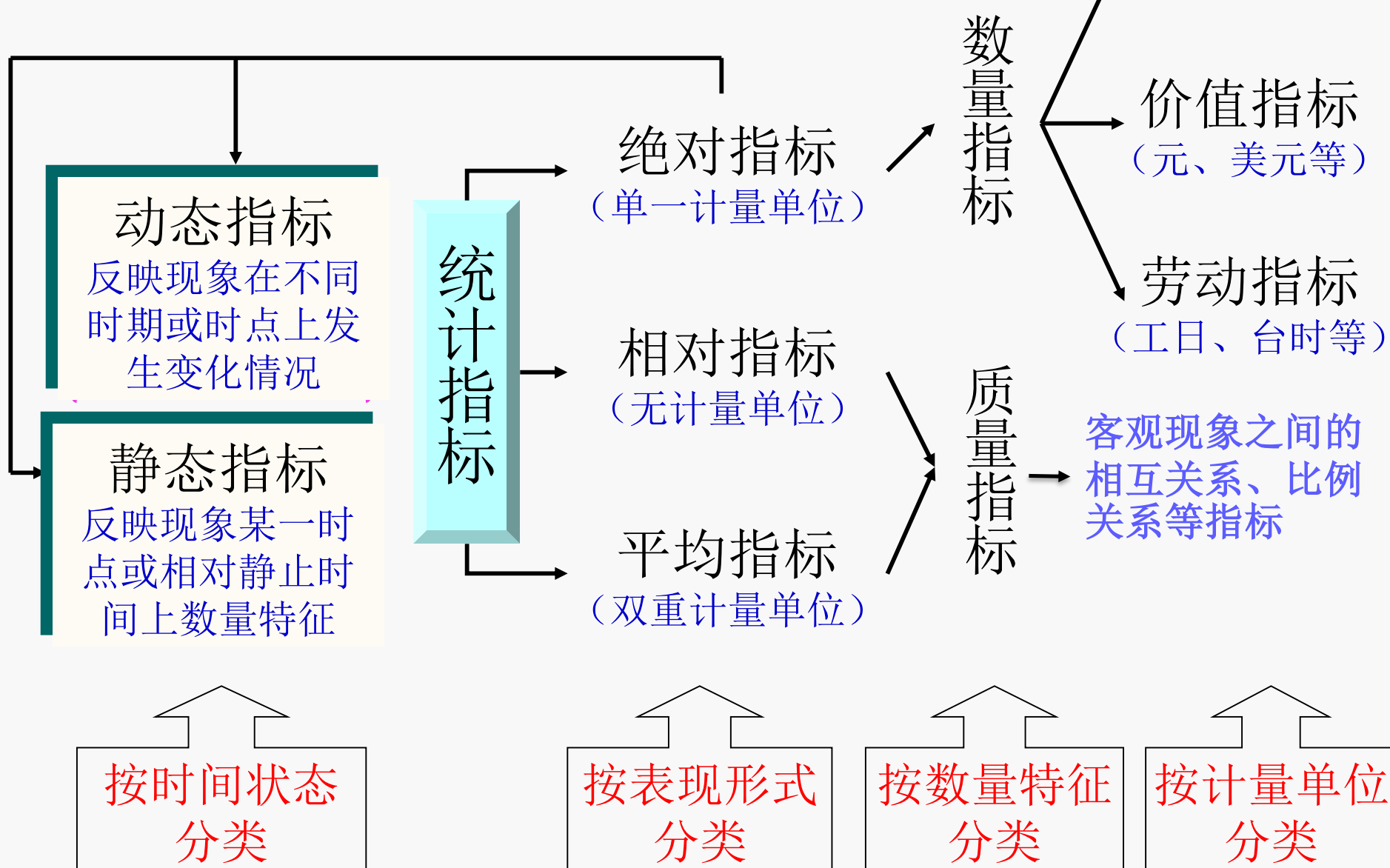
万人

特点： 数量性 具体性 综合性

注意统计口径



统计指标的基本分类



三、标志与统计指标

统计指标与标志的区别与联系

区别：

- **说明对象不同** 指标说明的是总体的特征，而标志则反映的是个体的特征
- **表现形式不同** 指标只能用数值来表现，而标志则既有用文字来表现的**品质标志**，也有用数值来表现的**数量标志**

联系：

- 统计指标的数值是根据说明个体特征的标志表现汇总得到的，既可以是对总体各单位标志值的综合，也可能是总体单位数的汇总



四、变量及其测量尺度

变量的种类

区分变量的意义

确定性变量

如何描述（规模、集中趋势等）

随机变量

如何推断（估计、检验等）

连续型变量

度量取得，理论上可以取任意小数，如人的身高、体重。

离散型变量

计数取得，只能表现为整数。



四、变量及其测量尺度

- **统计测量**：运用某种方法（**分类或标识**）使自然或社会经济现象量化，测量结果即形成**变量**。
- 统计测量尺度的种类：
 - 定类尺度
 - 定序尺度
 - 定距尺度
 - 定比尺度



四、变量及其测量尺度

定类尺度

- 按现象性质差异进行的辨别与区分，测量结果形成**定类变量或定类指标**
- 定类变量或指标确切的值是以**文字表述**的，可以用数值标识，但仅起标签作用
- 定类变量或指标的**各类别间是平等的**，没有高低、大小、优劣之分
- 定类变量只能进行是非判断，不能运算。

例如：性别、种族、运动项目等等



四、变量及其测量尺度

定序尺度

- 按现象顺序差异进行的辨别与区分，测量结果形成定序变量或定序指标
- 定序变量或指标确切的值以文字表述，也可以用数值标识，但仅起标签作用
- 定序变量或指标各类别间有高低优劣之分，不能随意排列
- 定序尺度的观测值既可以反映个体之间类别上的不同，也能够反映不同类别之间顺序上的差异。

例如：用一、二、三等品表示的产品等级变量



四、变量及其测量尺度

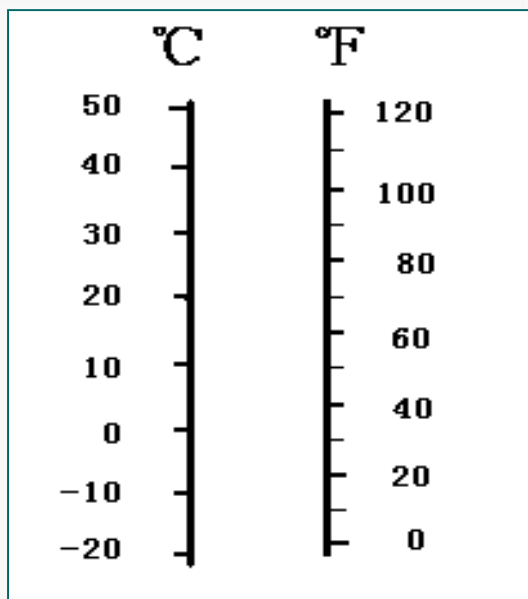
定距尺度

- 按现象绝对数量差异进行的辨别与区分，测量结果形成定距变量或定距指标
- 定距变量或指标的值以数字表述，有计量单位，可以进行加减运算
- 定距变量或指标各类别间数值有大小之分，但没有绝对的零点，不能乘除计算
- 定距尺度观测结果既能提供变量的分类信息和排序信息，也能提供变量分类和排序结果之间的距离信息

例如： 温度、纪年等



温度



天气预报：沈阳：最高温度 3°C ，最低 -7°C

大连：最高温度 6°C ，最低 -2°C

- ✓ 两地最高温度相差 3°C
- ✓ 沈阳最低温度较大连最低温度低 5°C
- ✗ 大连最高温度是沈阳最高温度的2倍

四、变量及其测量尺度

定比尺度

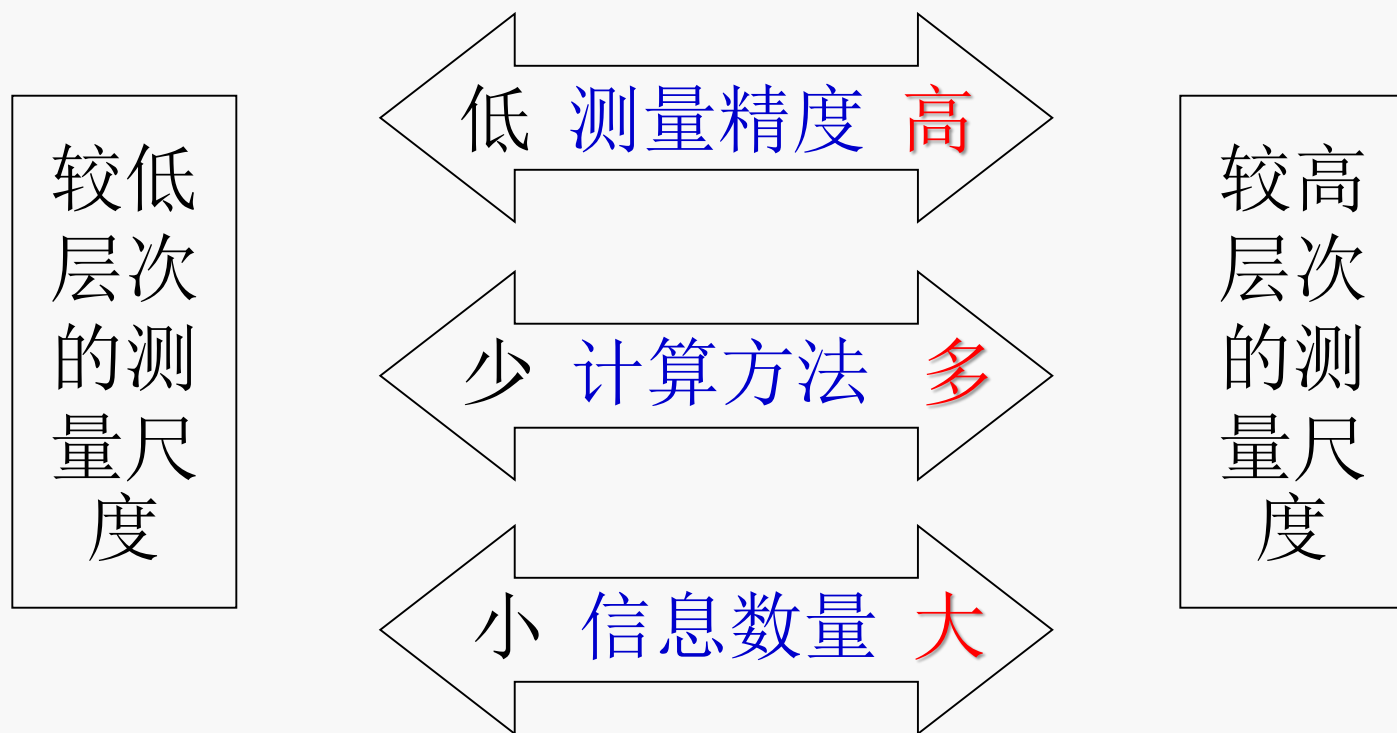
- 按现象绝对差异与相对差异进行的辨别与区分，测量结果形成定比变量或定比指标
- 定比变量或指标确切的值也以数字表述，有计量单位，可以进行加减运算
- 定比变量或指标有绝对意义上的零点，既可以加减运算，也可以乘除运算

例如：灯泡的使用寿命，人的年龄等



四、变量及其测量尺度

对测量尺度层次的判断



某甲、某乙的生命现象

测量精度

计算方法

信息数量

定类测量

甲、乙有生命

很低

不能计算，
只能判断
 $=, \neq$

甲、乙有生命

定序测量

甲为中年人，乙为少年人。

较低

$=, \neq, >, <$

甲、乙有生命
甲生命时间较乙长

定距测量

甲生于1961年、乙生于2006年。

较高

$=, \neq, >, <, +, -$

甲、乙有生命
甲生命时间较乙长
甲比乙大45岁

定比测量

甲60岁、乙15岁

很高

$=, \neq, >, <, +, -, \times, \div$

甲、乙有生命
甲生命时间较乙长
甲比乙大45岁
甲年龄为乙的4倍

常用统计分析软件中关于变量类型的定义：

测量尺度	SPSS的定义	SAS的定义
定类变量	Nominal	Nominal
定序变量	Ordinal	Nominal
定距变量	Scale	Interval
定比变量	Scale	Interval

第三节 SPSS及Excel基本操作

一、Excel基本操作

二、SPSS基本操作



一、Excel基本操作

数据的输入

- 数值、文本、日期或时间的输入
- 自动填充数据
 - 使用自动填充柄
 - 使用填充命令

Excel2010函数基础

- 公式的引入
- 函数的输入
 - 手动输入函数
 - 使用函数库输入



一、Excel基本操作

Excel2010函数基础

公式的引入：

“=” + “运算码”

值
单元格引用
单元格区域
区域名称
函数名

其间以一个或多个运算符连接，这些运算符包括算术运算符、逻辑运算符、字符运算符等



Excel中部分运算符含义及作用

名称	种类	含义	示例
算术运算符	+	加	1+2
	^	幂	4^2
	%	百分比	53%
比较运算符	=	等于	=B2/C3
	<	小于	B2<C3
	<=	小于等于	B2 <= C3
	<>	不等于	B2 <> C3
文本运算符	&	将多个文本连接起来产生连续的文本	“新年” & “快乐”，将产生“新年快乐”
引用运算符	,	联合运算符，将多个引用合并为一个引用	SUM（B2： C3， D4： E5） （引用B2： C3和D4： E5两个区域）

Excel公式中的运算符优先级顺序

运算符	名称	优先级顺序
:	区域	1
(空格)	交叉	2
,	联合	3
-	求反	4
%	百分数	5
^	幂	6
*和/	乘和除	7
+和-	加和减	8
&	连接	9
= ><>= <= <>	比较	10

一、Excel基本操作

数据分析工具

➤ 加载

选择“文件”→“选项”→“Excel选项”→“加载项”→“分析工具库”→“转到”，再选中“分析工具库”复选框→“确定”

➤ 调用

选择“数据”→“数据分析”，选择某种分析工具后，即可打开该分析工具对话框



二、SPSS基本操作



建立数据文件

- 直接输入数据
- 直接打开其它格式的数据文件
包括SPSS（.sav、.sys、.syd、.por）、Excel、Text等
- 使用数据库查询
- 导入文本文件



二、SPSS基本操作

定义数据的属性

——单击数据编辑窗口中的Variable View（变量视图）

- 变量名（Name）
- 变量类型（Variable Type）
- 变量宽度（Width）
- 小数点后位数（Decimals）
- 变量标签（Label）
- 值标签（Value Labels）
- 缺失值（Missing Values）
- 列宽（Column Width）
- 对齐格式（Align）
- 测度方式（Measure）



二、SPSS基本操作

保存数据文件

- 保存新文件：“File” → “Save”，弹出“Save Data As”对话框，在“Save as type”列表框中，选择对应的数据类型 → 在文件名文本框内输入新文件名 → “Save”；
- 再保存已保存过的数据：选择“File” → “Save”；
- 另存为：“File” → “Save as”，在弹出“Save Data As”对话框中，输入新文件名。

导出数据文件

- 选择“File” → “Export to Database”，弹出“Export to Database Wizard”对话框，选中所需的数据库类型，根据向导的提示即可将数据导出到指定的数据库



