

卫生部“十一五”规划教材

《医学统计学》 主编：马斌荣

第一章

医学统计中的基本概念

(**Basic Concept in Medical Statistics**)

武汉大学流行病与卫生统计学系 宇传

周次	日期	星期	节次	授课内容	任课老师	职称	授课地点
1	9-8	四	1-3	医学统计中的基本概念	宇传华	教授	211
2	9-15	四	1-3	集中趋势的统计描述	胡樱	副教授	211
3	9-22	四	1-3	实习一 SPSS简介	胡樱	副教授	5号楼机房
4	9-29	四	1-3	变异程度的统计描述	胡樱	副教授	211
5	10-6	四	1-3	抽样误差与假设检验	胡樱	副教授	211
6	10-13	四	1-3	实习二 基本统计描述	胡樱	副教授	5号楼机房
7	10-20	四	1-3	t检验	胡樱	副教授	211
8	10-27	四	1-3	方差分析	胡樱	副教授	211
9	11-3	四	1-3	实习三 t检验、方差分析	胡樱	副教授	5号楼机房
10	11-10	四	1-3	相对数及应用	胡樱	副教授	211
11	11-17	四	1-3	χ^2 检验	胡樱	副教授	211
12	11-24	四	1-3	实习四 卡方检验	胡樱	副教授	5号楼机房
13	12-1	四	1-3	非参数检验	孙静	讲师	211
14	12-8	四	1-3	线性相关与回归	孙静	讲师	211
15	12-15	四	1-3	实习五 非参数检验	孙静	讲师	5号楼机房
16	12-22	四	1-3	统计表与统计图	孙静	讲师	211
17	12-29	四	1-3	医学实验设计与诊断试验的评价	陈冬峨	教授	211
18	1-5	四	1-3	实习六 非参数检验相关与回归分	陈冬峨	教授	5号楼机房

第一节 概述

统计学(**statistics**)

统计学是关于数据 (**data**) 的科学，是从数据中提取信息的一门学科，包括设计、搜集、整理、分析和表达等步骤。

根据研究领域和研究对象，统计学又分为：数理统计、经济统计、生物统计、医学统计、卫生统计……

医学统计学 (medical statistics)：用统计学的原理和方法研究医学中的问题。

卫生统计学 (health statistics)：与医学统计学基本相同，但更侧重于社会、人群的卫生问题。

表 1 两种教学效果评分比较

分, $\bar{x} \pm s$

项 目	传统式教学法	模仿式教学法
基本知识	7.91±1.47	7.80±1.32
兴趣性	7.41±1.32	7.96±1.35*
主动性	6.95±1.32	8.17±1.34*
合理性	7.86±1.18	8.06±1.22
联系实际	7.59±1.14	7.32±1.42
自学能力	6.85±1.56	8.18±1.33*
归纳能力	7.00±1.35	7.80±1.28*
思维能力	6.91±1.57	7.66±1.33*
综合能力	7.34±1.35	7.82±1.24*

与传统式教学法比较 * $P < 0.01$

北京某医院某大夫使用“乌贝散”，治疗胃溃疡病出血107例，有效101例，有效率为94.4%。

别的医院，其他大夫使用“乌贝散”来治疗胃溃疡病出血，其有效率也一定是94.4%吗？

求 95% 置信区间 (confidence interval)

某医院有位老大夫，用“冠心灵”治疗冠心病，其对照组用西药，观察结果如下表：

表1-1 冠心灵与单纯西药疗效对比

	显 效	有 效	无 效	合计
单纯西药	9	25	6	40
冠 心 灵	19	18	5	42

冠心灵是否比单纯西药有效？

假设检验 (test of significance)，

俗称要算 P 值。

为什么要学统计学？

1. 采用统计学方法，发现不确定现象背后隐藏的规律。

变异（variation）是社会和生物医学中的普遍现象。

变异使得实验或观察的结果具有**不确定性**，如每个人的身高、体重、血压等各有不同。

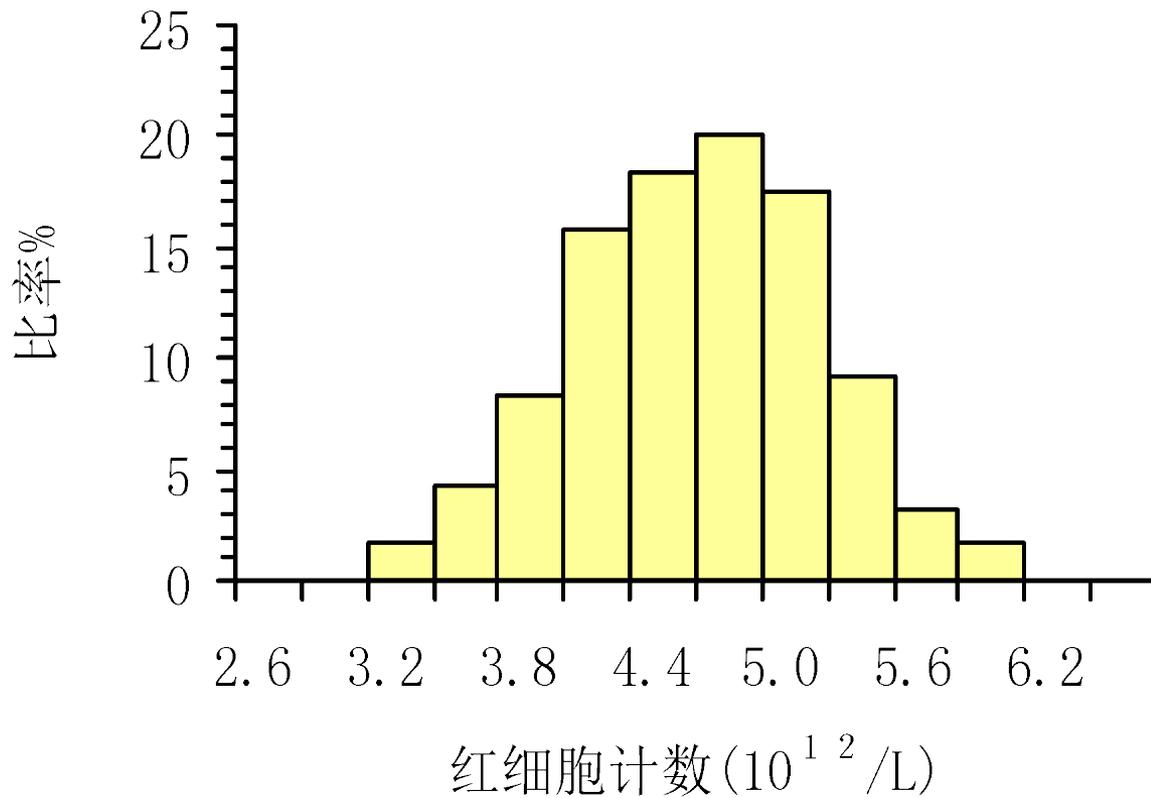


图 1-1 120名正常成年男子
红细胞计数直方图

最大值=6.18，最小值=3.29，极差=2.89。

算术均数=4.72，标准差=0.57。

2. 用统计学思维方式考虑有关医学研究中的问题

“阳性”结果是否是虚假联系？

某感冒药治疗**1**周后，治愈率为**90%**，能否说该感冒药十分有效？

“阴性”结果是否是样本含量不足？

有人曾对发表在**Lancet, N Engl J Med, JAMA**等著名医学杂志上的**71**篇阴性结果的论文作过分析，发现其中有**62**篇（**93%**）可能是由于**样本含量**不足造成的假阴性。

3. 保证你的研究论文能通过统计学审查

据国外**60~80**年代对不同医学杂志发表论文的调查，有**统计错误的论文的百分比最高达72%**，最低也有**20%**。

国内**1984**年对《中华医学杂志》、《中华内科杂志》、《中华外科杂志》、《中华妇产科杂志》、《中华儿科杂志》**595**篇论文的调查结果，相对数误用为**11.2%**，抽样方法误用**15.9%**，统计图表误用**11.7%**。某研究者**1996**年对**4586**篇论文统计（中华医学会系列杂志仅占**6.9%**），数据分析方法误用达**55.7%**。

JAMA，新英格兰医学杂志（**NEJM**）、英国医学杂志、中华医学会杂志对来稿都有统计学要求或统计学指导。

国际生物医学杂志编辑协会在其《生物医学期刊投稿的统一要求》中也包含了统计学要求。

Uniform Requirements for Manuscripts
Submitted to Biomedical Journals

<http://>

4. 获得循证医学证据的主要手段

“良好愿望的医学”（**well-meaning medicine**）转入“以证据为基础的医学”（**evidence-based medicine, EBM**）需要有统计学方法的支持。

全世界的医学期刊每年大约刊登600万学术论文，但能作为可靠“证据”的论文并不多。

第二节 医学统计资料的来源与分类

一、原始统计数据的来源

1. 常规保存数据

医院信息系统（**HIS**）、统计报表等

2. 现场调查记录

3. 实验记录

4. 其他

二、统计数据分类

定量数据

只有认识了数据的特点，才能正确地选用统计分析方法

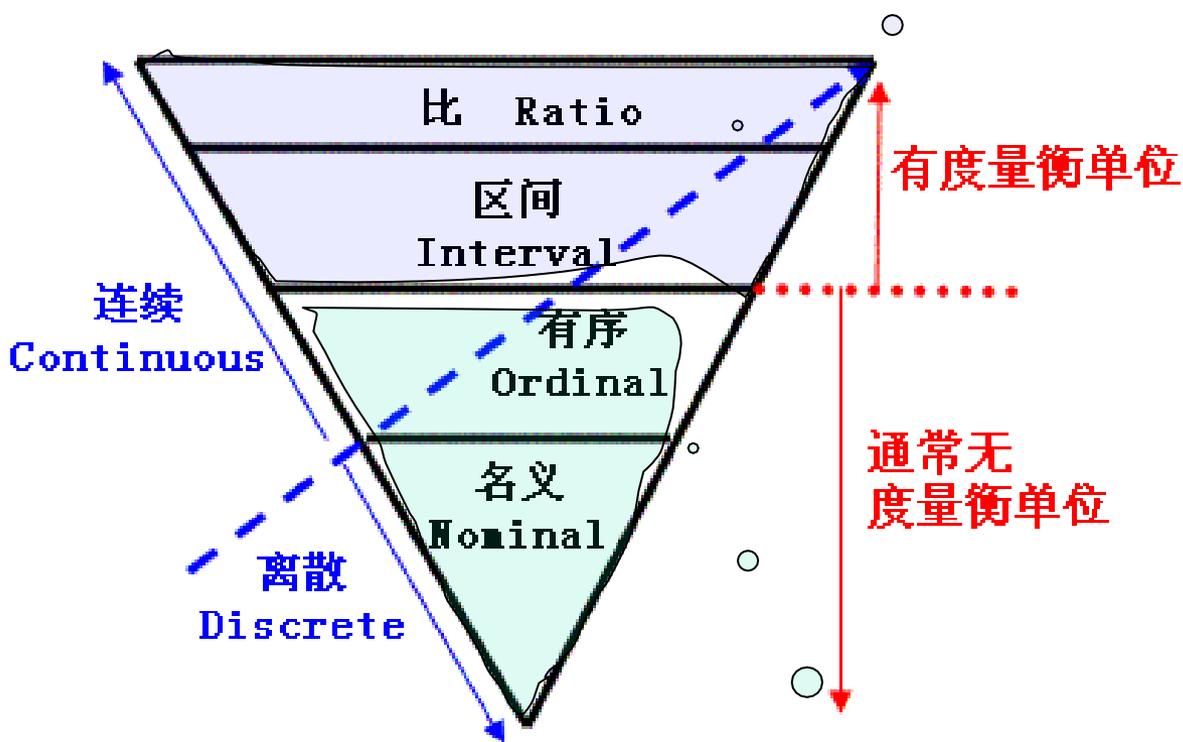


图1 数据分类

按变量测量的精确程度由低到高，将数据分类为：名义数据（如性别、婚姻状况）、有序变量（如疗效，类别间差别大小难以度量）、区间变量（如摄氏体温，类别间差别有实际意义）、比变量（如身高，除具有区间变量的特征外，还具有真实意义的零点。摄氏温度的零点为水结冰时温度，并非绝对意义的零点，所以它不属于比变量）

定性数据

（一） 定性与定量数据的别名

1、 定性数据（**qualitative data**）

（二项与多项）分类数据(**categorical data**)

计数数据（**enumeration (counting) data**）

2、 定量数据（**quantitative data**）

计量数据（**measurement data**）

区间数据（**interval data**）

数值数据（**numerical data**）

3、 等级资料（**ranked data**）

有序资料（**ordinal data**）

(二) 三类数据间的关系

例：一组20~40岁成年人的血压

等级数据	<8	低血压	定量数据
	8~	正常血压	
	12~	轻度高血压	定性数据
	15~	中度高血压	
	17~	重度高血压	

以12kPa为界分为正常与异常两组，统计每组例数

观察单位
observations

Units; elements

变量
variables

个体 individuals

住院号	年龄	身高	体重	住院天数	职业	文化程度	分娩方式	妊娠结局
	27	165	71.5	5	无	中学	顺产	足月
	22	160	74.0	5	无	小学	助产	足月
	25	158	68.0	6	管理员	大学	顺产	足月
	23	161	69.0	5	无	中学	剖宫产	足月
	25	159	62.0	11	商业	中学	剖宫产	足月
	27	157	68.0	2	无	小学	顺产	早产
	20	158	66.0	4	无	中学	助产	早产
	24	158	70.5	3	无	中学	助产	足月
	29	154	57.0	7	干部	中学	剖宫产	足月

Quantitative data
定量资料

Qualitative data
定性数据

(三) 三类数据的编码与计算

(1)

例：一组20~40岁成年人的血压

等级数据	<8	低血压	4	1	定量数据
	8~	正常血压	10	2	
	12~	轻度高血压	13	3	定性数据
	15~	中度高血压	16	4	
	17~	重度高血压	18	5	

以12kPa为界分为正常 (=0) 与异常 (=1) 两组，统计每组例数

(三) 三类数据的编码与计算

(2)

1. 定量数据: 允许计算均数、标准差等 (可采用 t 、 F 检验等, 可当做有序或定性数据处理)
2. 定性数据: 编码是任意的, 不能对编码执行均数标准差等计算, 但可计算率或比 (可采用卡方检验等)

调查14名男子的婚姻状况

分类	编码	频数 (frequency)
单身	1	3
已婚	2	5
离异	3	2
分居	4	

3. 有序数据: 允许基于顺序的计算, 如计算中位数、百分位数 (可当做定性数据处理)

第三节 统计学常用的基本概念

一、变量与随机变量

Variable and random variable

- 1、变量——可以测量的任何特征或属性**
Any characteristic or attribute that can be measured.（不同个体结果可能不同）
- 2. 随机变量——在概率论中称变量为随机变量**

根据观察数据之间有无缝隙（gap），常将数据分类为离散型（discrete）变量（有缝隙）与连续型（continuous）变量（无缝隙）两大类。

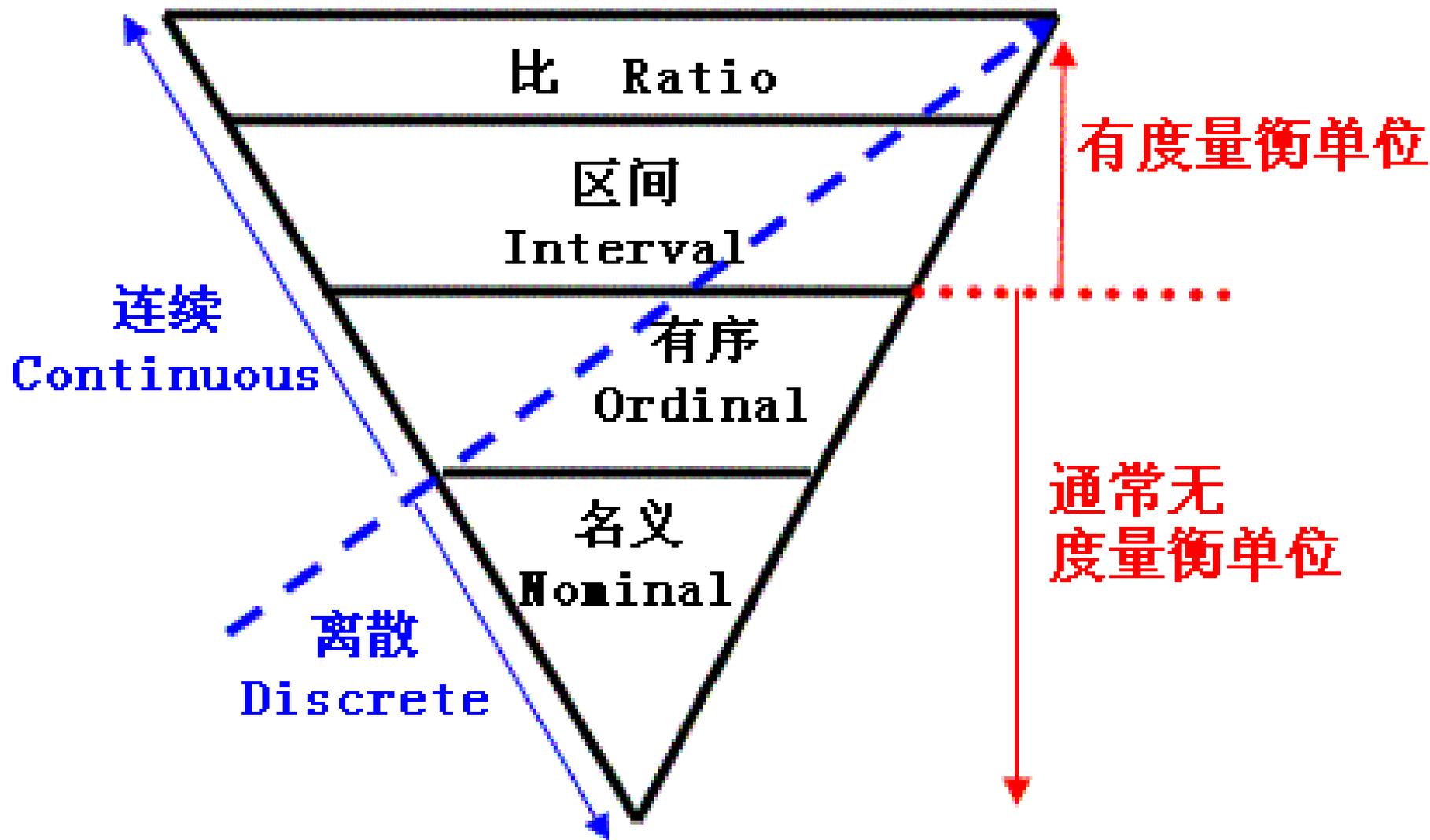


图1 数据分类

二、测量和测量尺度

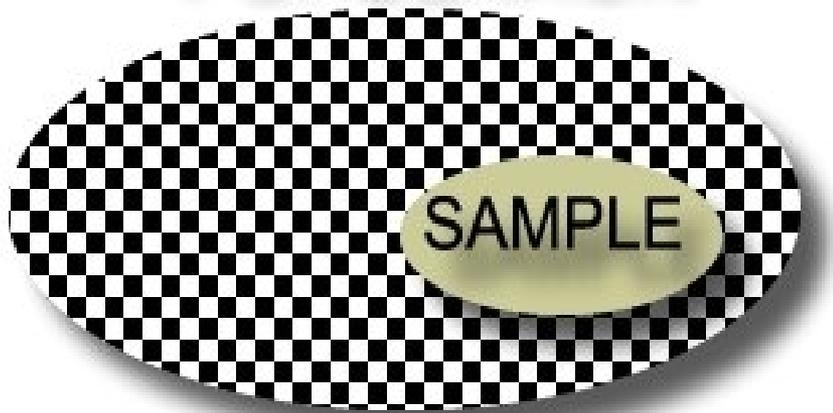
对随机变量的取值过程为测量。取值所采用的标准为测量尺度。

采用统一的标准化的尺度进行测量，其测量结果一般是稳定的和一致的，例如身高、胸围等形态指标的测量均采用统一的测量尺度(全国统一研制的身高计、胸围尺均以厘米为记录单位，体重以公斤为记录单位，肺活量以毫升为单位等)

三、总体与样本

population and sample

POPULATION



总体：根据研究目的确定的**同质**研究对象的**全体**（集合）。分有限总体与无限总体

样本：从总体中随机抽取的部分研究对象

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/545320001001012010>