

课程安排

- 48学时
- 1~12周（理论课），每周3学时
- 13~15周（理论课），国庆、节假日等课时自动删减
(上其他班听课、自学、请老师补课)
- 考试2学时



学习要求

1. 考试卷面成绩70-80分
2. 自备作业本，实验报告20-30分
3. 每次下课前交实验报告，否则，迟交一天扣10分（二天扣20分……）
4. 自备计算器（带统计功能），如有兴趣，可自学**SPSS16、SAS、R**。



2014年下学期统计学教学安排

课程名称	专业班级	人数	总学时	周学时	上课时间	地点	任课教师
					星期(节)	教室	
卫生统计学	11级预防	38	80	9	一(5.6.7) 五(5.6.7) 三(1.2.3.4.5)	203 1 实验	易尚辉 邹享玉 吕媛
医药统计学	12级医学检验	47	54	3	三(5.6.7)	104	邹享玉
医学统计学	12级临床(一班)	74	48	3	二(1.2.3)	2	易尚辉
医学统计学	12级临床(二班)	82	48	3	一(1.2.3)	505	邹享玉
医学统计学	13级护理学	33	48	3	四(1.2.3)	203	吕媛
医学统计学	13级护理(树达)	81	36	3	二(5.6.7)	403	易尚辉
医学统计学	12级临床(树达一班)	89	36	3	五(1.2.3)	3	查文婷
医学统计学	12级临床(树达二班)	88	36	3	四(5.6.7)	4	查文婷



“非常痛心地看着，因为数据分析的缺陷和错误，那么多好的生物研究工作面临着被葬送的危险”。

— F. Yates, M.J.R. Healy

为什么要学医学统计学？

统计知识的运用

撰写论文：报告自己观察或实验的研究结果。

阅读论文：吸收新知识，了解学术进展。



1996年，有机构对申报科技成果的**4586篇**科研论文分析，统计方法使用率为**76%**。

医学论文中统计运用错误，除了影响论文的科学性，还可能导致严重的伦理学问题。

60年代到80年代，国外医学杂志调查表明：**20%~72%**的论文有统计错误。

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：

相对数误用占 **11.2%**，抽样方法误用占**15.9%**，统计图表误用占**11.7%**

1996年对**4586篇**论文统计（中华医学会系列杂志占**6.9%**），数据分析方法误用达**55.7%**。

2001年《中华预防医学杂志》：中华医学会系列杂志误用约**54%**（1995）。



英国统计学家R.A.Fisher (1890-1962) 对遗传学家Mendel (1822-1884) 杂交试验结果的评价:

Mendel的豌豆杂交试验:



杂交试验结果:绿色种子的频率2001/8023

期望值= $8023 \times 0.25 = 2006$, 标准差=39

$\text{Prop}(2001.5 \sim 2011.5) = 0.11$, 结合Mendel历次报告的其它试验结果,每次都有如此好的吻合的概率约10万分之4。



为什么要学医学统计学?



第一章 绪论

- 第一节 统计学的几个基本概念
- 第二节 医学统计工作的基本步骤
- 第三节 统计学发展简史





统计学(Statistics)

- ▶ 是一门研究数据的搜集、整理、分析的科学。





医学统计学 (medical statistics)

定义: 是研究医学数据的搜集、整理、分析的科学。

特点: 1、资料的数字化, 数量反映质量
2、由偶然性 (不确定性) 的剖析中, 发现事物的必然性 (确定性)



医学统计学

学习内容:

- 1、统计学的基本原理、方法——基础（学习重点）
- 2、业务统计——应用（如：临床医学、预防医学、卫生事业管理）



医学统计学

目的：医学科研的数学工具

核心：计算和比较反映群体综合数量特征的统计指标。

例如：期望寿命——反映人群健康状况的指标

婴儿死亡率——反映卫生服务质量的指标





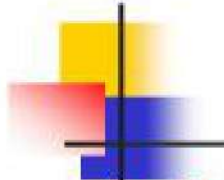
第一节 统计学的几个基本概念





一、个体与变量、总体与样本





个体（观察单位等）：统计研究中的基本
单位——据研究目的而定

变量（观察指标等）：要研究的个体特征
例如：身高、体重、性别、血型、反
应、疗效等





变量值：变量的取值

例如：身高 1.65米 体重 52公斤

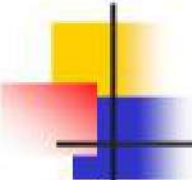
性别 女 血型 “O”型

反应 阴性 疗效 好转

资料：由多个变量值构成

例如：血糖的样本资料





同质：给个体规律的一些相同性质（使研究变量的已知影响因素齐同）

如：变量为身高，长沙市2004年7岁正常男童即为同质

变异：同质个体的变量值的差异

如：长沙市2004年7岁正常男童的身高值各不相同

特点：生物个体变异较大（原因：难于掌握的因素造成。如：影响身高的有遗传、营养等因素）



总体 (population) :

定义: 根据研究目的确定的同质个体 (变量值) 的全体。

例如: 长沙市2004年7岁正常男童 (身高值) 的全体

有限总体 (finite population): 明确规定了空间、时间、人群范围内有限个观察单位的总体。

无限总体 (infinite population): 没有时间和空间范围的限制, 其观察单位的全体数只是理论上存在, 因而可视为“无限”的总体。





样本 (sample)

定义：从总体中随机抽得的部分观察单位，
其实测值的集合

例如：长沙市2002年7岁正常男童中随机抽取
200名，其身高值构成一个样本



抽样研究

样本信息

推断

总体特征

(样本统计指标、统计量)

(总体统计指标、参数)

注意：获取样本信息是**手段**，推断总体特征是**目的**。

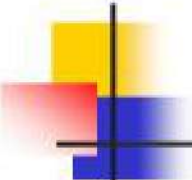




二、变量与资料

变量分类是资料分类的基础






计量资料：（measurement data）又称定量资料或数值变量资料。为测定每个观察单位某项指标的大小而获得的资料。其变量值是定量的，表现为数值大小，一般有度量衡单位。

例如：身高资料、体重资料





计数资料 (enumeration data) 又称定性资料或无序分类变量资料。为将观察单位按某种属性或类别分组计数，分组汇总各组观察单位数后而得到的资料。其变量值是定性的，表现为互不相容的属性或类别。

- ▶ **二分类**：如性别、生死、疾病有无；
- ▶ **多分类**：如A、B、O、AB血型。



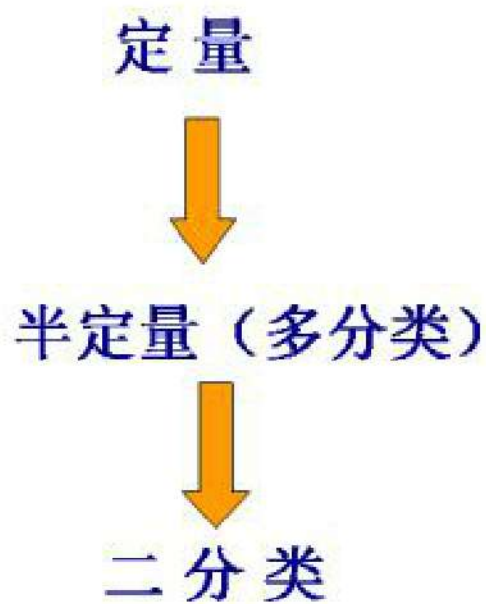
等级资料 (ranked data) 又称半定量资料或有序分类变量资料。为将观察单位按某种属性的不同程度分成等级后分组计数，分类汇总各组观察单位数后而得到的资料。其变量值具有半定量性质，表现为等级大小或属性程度。

例如：观察用某药治疗某病患者的疗效，以每名患者为观察单位，结果可分为治愈、显效、好转、无效四级。

统计分析方法的选用，是与资料类型密切联系的。



资料分类可互相转换



例如：测得5人的WBC（个/ m^3 ）数如下：

3000	6000	5000	8000	12000	数值变量
过低	正常	正常	正常	异常	分类变量

- ▶ 若按正常3人，异常2人分组→二分类变量
- ▶ 若按过低1人，正常3人，过高1人分组→等级资料





三、误差





定义：实测值与真值之差。

1、**随机误差：**不恒定的、随机变化的误差，由多种尚无法控制的因素引起。无方向性。

主要指重复测量产生的测量误差和抽样过程产生的抽样误差。

通常，测量误差远小于抽样误差，因此统计学主要考虑**抽样误差**。





非随机误差又可分为系统误差和非系统误差两类：

2、**系统误差**：实验过程中产生的误差，它的值或恒定不变，或遵循一定的变化规律，其产生原因往往是可知的或可能掌握的，大小变化有方向性。

3、**非系统误差（过失误差）**：研究者偶然失误而造成的误差。





四、频率与概率





1. **频率** (frequency) , 假设在相同条件下, 独立地重复做 n 次试验, A 在 n 次试验中出现了 m 次, 则比值 m/n 称为随机事件 A 在 n 次试验中出现的频率。当试验重复很多次时, 有

$$P(A) \approx m/n$$



2. 概率 (probability) : 概率是度量随

机事件发生可能性大小的一个数值。

设在相同条件下，独立地重复 n 次试验，随机事件 A 出现 f 次，则称 f/n 为随机事件 A 出现的频率。当 n 逐渐增大时，频率 f/n 趋向于一个常数，则称该常数为随机事件 A 的概率，可记为 $P(A)$ ，简记为 P 。



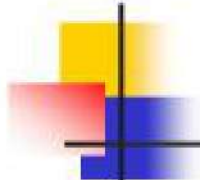


取值: $0 \leq P(A) \leq 1$

若 $P(A)=1$, A 为必然发生事件 }
若 $P(A)=0$, A 为必然不发生事件 } 非随机事件;

若 $0 < P(A) < 1$, A 为随机事件。



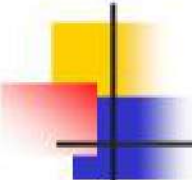


频率是就样本而言的，而概率从总

体的意义上说的， m/n 是概率 $P(A)$ 的估计

值。试验次数越多，估计越可靠。





小概率事件: 统计分析中的很多结论都基于一定置信程度下的概率推断, 习惯上将 $P(A) \leq 0.05$ 或 $P(A) \leq 0.01$ 称为小概率事件, 表示在一次实验或观察中该事件发生的可能性很小, 可视为很可能不发生。





第二节 医学统计工 作的基本步骤





一、设计

主要指统计设计，是影响研究能否成功的最关键环节，是提高观察或实验质量的重要保证。

内容：包括对资料搜集、整理和分析全过程的设想与安排。





实验设计的三个基本原则

1. 随机化 (randomized sampling)
2. 重复 (repletion)
3. 对照 (control)





二、搜集资料

目的：指应采取措施使能取得准确可靠的原始数据。

资料来源：

- ① 统计报表。
- ② 经常性工作记录。
- ③ 专题调查或实验研究。
- ④ 统计年鉴和统计数据专辑。

要求：

- 1、随机性
- 2、样本含量足够大





三、整理资料

是将原始数据净化，系统化和条理化，以便为下一步计算和分析打好基础的过程。

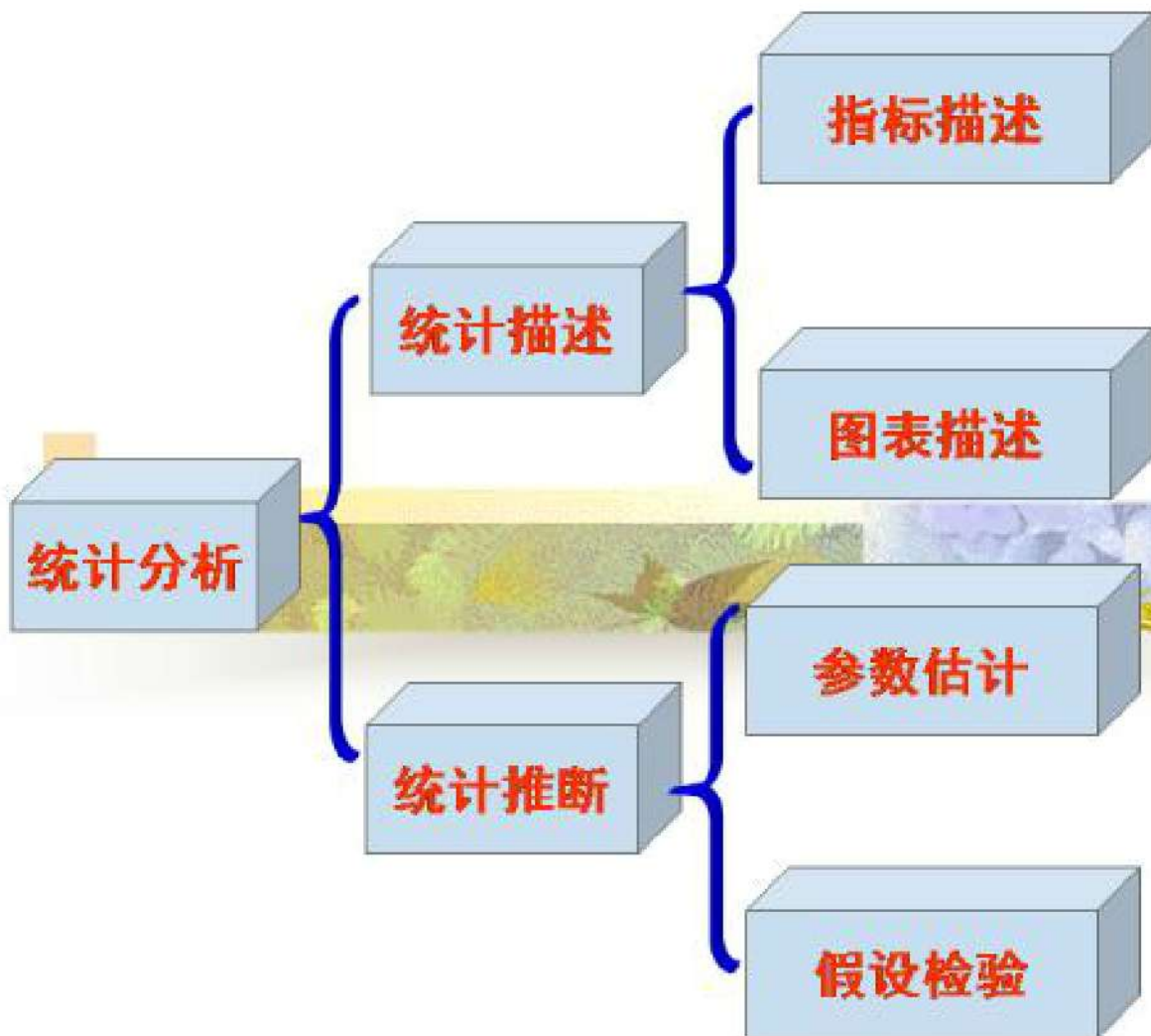


四. 分析资料 (统计分析)

目的：在表达数据特征的基础上，阐明事物的内在联系和规律性。包括两方面：

- 1. 统计描述**：用统计指标、统计图表对资料的数量特征及分布规律进行测定和描述。
- 2. 统计推断**：用样本信息推断总体特征：①参数估计，②假设检验。







第三节 统计学发展简史





I. 早期

公元前3050年古埃及人为修建金字塔筹集建筑费，对全国的人口和财产进行了普查。





2. 近代

19世纪，应用数学家为解决赌徒们在博彩中出现的输赢概率问题逐渐形成和发展了概率论，从而为统计学的发展奠定了坚实的基础。





3. 现代

计算机和统计软件如SAS、SPSS的出现——使统计学得到了突飞猛进的发展。





生物统计学著名

历史人物其人其事





1. 数学神童

Gauss (1777 – 1855)





德国数学神童**高斯**在孩童时就能够

快速计算出“ $1+2+3+\dots+100=5050$ ”的答案。

他对统计的主要贡献除了正态分布（

normal distribution）外，还有算术均数、几

何均数、二项分布理论和最小二乘法等。





其实，在Gauss之前就有人发

现了正态分布，但只有他很快将之应用于天文学研究，并使其应用价值广为人知，故正态分布又称高斯分布(Gaussian distribution)。





2. 从酿酒师到统计

主管的 *Gosset* (1876-1937)





1899年Gosset在英国都柏林一酿酒

公司做酿酒师，在对小样本进行质量控制的 t 分布的研究中发现了 t 分布，其论文1908年以Student为笔名发表。自此开创了小样本统计的新纪元。





非常有趣的现象是，Gosset的数学欠佳，因而不能自己解决 t 检验的理论和应用问题，其最后的完善，是由Fisher, Neyman 和E Pearson 先后完成的。正如后人评价的那样：“Gosset提出实际问题，Fisher 和 E Pearson 将其转成统计问题，Neyman用数学解决问题”。





大家熟知的公式:

$$t = (\bar{X} - \mu) / (S / \sqrt{n})$$

实际上是1925年Fisher在《 t 分布的应用》一文中定义的。

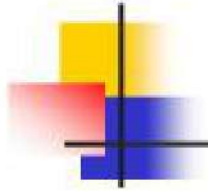




3. 医院统计与卫生管理统计的

先驱——*F. Nightingale* (1820-1910)





1854年9月南丁格尔出任土耳其英国总医院的女护士总管。鉴于医院没有任何医疗统计，登记也十分简单，仅仅记载下某年某日一个人死亡了。因而，她紧急提议建立医院的**医疗统计室**。她认为**图表**是她发明的一种表达形式。





1857年底，按照南丁格尔的建议在陆军部建立了四个下属委员会，陆军部统计委员会是其中之一。1858年她当选英国皇家统计学会会员，是该学会最早的女会员。1860年，第四届国际统计大会议题之一就是“南丁格尔小姐的医院统计标准化计划”，南丁格尔在会上报告了她的论文“医院统计学”。1862年，维多利亚出版社出版了她的著作《医院统计与医院规划》。






4. 把统计作为“披荆斩棘”

工具的人类学家——*Galton*





“回归”一词来源于Galton的《自然遗传》。他是达尔文的表兄弟，人类学家，早年学医，曾在剑桥大学念书。尽管他的数学不是很好，但在人类学和优生学研究中萌发的统计学思想，对生物统计的发展产生了深远影响，如“回归”、双变量正态分布的概念等。





当人类科学的探索者在问题的丛林

中遇到难以逾越的障碍时，唯有统计学

工具可以为其开辟一条前进的通道。

— *F Galton* (1822-1911)





学习方法

- 掌握基本概念
- 重在正确应用

- 选择恰当方法
- 满足应用条件
- 善于解释结果



教学目的

培养统计思维

正态分布

均数抽样分布

假设检验基本思想

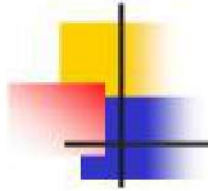
用于医学科研





欢迎报考本专业
研究生、博士





谢谢大家!

