

常用医学统计



湖南医学院卫生系统计学教研组

目 录

第一章 统计资料的收集和整理

第一节 统计资料的收集

第二节 统计资料的整理

第二章 相对数

第三章 均数与标准差

第一节 均数

第二节 标准差

第四章 抽样问题

第一节 抽样误差

第二节 显著性测验

第三节 卡方(χ^2)测验

第五章 统计表和统计图

第一节 统计表

第二节 统计图

练习题



常用医学统计方法

概 述

伟大领袖毛主席教导我们：“胸中有‘数’”。意思是说，对情况和问题一定要注意到它们的数量方面，要有基本的数量的分析。任何质量都表现为一定的数量，没有数量也就没有质量。我们有许多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计，基本的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无‘数’，结果就不能不犯错误。”

在统计工作中，也曾发生过两条路线的激烈斗争。由于刘少奇、林彪，特别是“四人帮”的干扰和破坏，资产阶级学术权威鼓吹“统计万能论”，把医学统计描绘“形式主义”、“机械哲学”，严重脱离政治，脱离群众，脱离实际，继而又散布“统计无用论”，破坏统计又作为无产阶级政治服务。在毛主席亲自发动和领导的无产阶级文化大革命中，批判了刘少奇、林彪，特别是“四人帮”在医学统计工作上的流毒，使医学统计工作能够在正确路线指引下前进，更好地为社会主义革命和社会主义建设事业服务。

胸中有“数”与胸中无“数”，只是一字之差，但是却反映了两种不同的思想和两种不同的作风。要不要统计，是不是胸中有“数”，是关系到坚持唯物主义的反映论，还是搞唯心

主义的先验论的原则问题。常用医学统计方法就是研究人民疾病和健康状况的数值特征的重要手段。如常用于反映人民健康状况的数值特征有：出生率、死亡率、发病率等；常用于反映疾病数值特征有：潜伏期、病程、传染、治愈率等。这些数值都是经过调查研究得来的。用这些统计数字（即统计指标）能够说明人民健康水平和医疗卫生工作状况，进行有关疾病和人民健康状况的研究。

根据医学专业的实际需要，本篇着重介绍统计资料的收集与整理、相对数、平均数与标准差，抽样问题及统计表格与统计图等。学员学习时，要注意理论密切联系实际，搞清概念，加强实践，培养严谨的工作作风，发扬实事求是的科学态度。同时运用马克思主义的观点批判可能存在的各种错误观点。

第一章 统计资料的收集和整理

统计的基本步骤可分成：^①统计资料的收集、^②整理和^③分析等三步。这些步骤之间是彼此联系和互相制约的，每个步骤的正确与否可以直接影响统计工作的质量。

第一章 统计资料的收集

一 统计资料收集的要求

统计资料的收集是进行统计研究的第一步，任何原始资料都应按计划加以收集。原始资料的缺陷，会造成统计上难以补救的损失。例如原始资料的数目不足，项目不全，记录不正确以及缺乏代表性和对比条件等，常可使结果不能说明问题。为了正确收集原始资料，必须注意以下几点要求：

(一) 资料的完整、正确和及时

完整：原始资料必须具有完整性，要求填写的项目都要填写完整，在数字上没有重复或遗漏。例如传染病报告卡发出重复或遗漏时，便不能真实地反映传染病的流行规律。

正确：原始资料的收集和登记，应该严格按照要求填写。各项目的界限应有明确规定。例如出院卡中的治愈、好转、无变化、死亡等均应按说明加以正确区分，否则这些数字便失去意义。

及时：统计资料往往是反映某一特定时间、地点、条件下的具体情况。例如各种传染病报告，便有严格的时间性，如果不能及时填写报告，便失去防治时机，对工作带来不利的影响。

(二) 资料的数目、对比条件和随机分配

数量：由于事物存在着个体差异，因此观察例数必须有相当数量，才能反映了物的客观规律，特别是在制订各种生理正常值或研究发病率或病死率较低的某些疾病的流行规律时，更应该有较多的观察例数。在医学科研的实物实验中，由于各种条件控制较好，观察例数可以相对减少。

对比条件：在医学研究中，经常将观察对象分成实验组和对照组来进行对比。对比时，应尽可能使对比的各组，除了实验条件不同以外，其他可能对结果有影响的因素都尽可能一致。例如进行不同药物的疗效观察时，如果两组病人的病情不同，常可使结果产生偏性，甚至于得出错误的结论。

随机分配：为了取得对比条件的一致，在分组时应力求做到“随机”分配。所谓随机分配就是使每一个观察对象都有同等的机会被分配到各组中去，避免研究者凭主观意愿而加以选择。常用最简单的随机分配方法是抽签法。

二、统计资料收集的方法

收集原始资料首先要制订计划，现将调查计划和实验设计的主要内容分述如下。

(一) 调查计划

调查计划应包括整个调查研究过程中各个阶段的总体要求和安排。调查计划是否周密、妥善，将会直接影响调查的结果。

制订调查计划应考虑以下内容。

1. 目的。调查目的就是调查所要解决的问题。调查目的应力求明确和具体。例如调查某地居民肠寄生虫病的感染情况，便应该明确是只了解肠寄生虫病的感染率呢？还是在深入了解了肠寄生虫病感染与地区、性别、年龄、卫生习惯等的关系呢？因为调查目的不同，所调查的项目、范围、方式以及组织也就不同。漏掉了应该调查的项目，或调查了一些不必要的项，都会浪费人力物力，而得不到预期的结果。

2. 对象和单位。根据调查目的来规定调查对象。例如调查某地居民中肠寄生虫病感染情况，则该地全体居民都应是调查对象，而每一个居民就是一个调查单位。又如调查某种药物对治疗青光眼的效果，则那些使用该药治疗的青光眼患者便是调查对象，而每天服用该药治疗的眼睛便是调查单位。

3. 范围。调查范围可分为“普查”和“抽查”两种，这是根据调查的目的和任务而定的。例如调查者在立即进行普查工作时，便需在调查中摸透全部治疗对象，这就需普查才行。如果只是为了说明该病的流行情况，那么只要进行“抽查”就可以达到目的。由于“抽查”的范围较小，可以使调查内容比较深入和细致；但是抽取样本必须遵守随机化的原则，才能利用“样本”来推知“总体”。

4. 方式。常用调查方式有直接观察、^①询问和^②报表三种。例如直接由调查者进行探查而收集的资料为直接观察法，这种

方式比较准确可靠。单纯依靠询问调查对象而记录的为询问法。另外，由调查对象自己照调查表的项目填写答案的为填表法。询问法和填表法虽不如直接观察法准确可信，但可以节省人力物力和时间。以上三种方式可根据具体情况加以选用。

5. 项目：就是调查的内容。调查项目是根据调查目的和所要分析的问题而提出的。例如调查某地居民肠寄生虫病感染情况，便需把该地应调查的居民姓名、年龄、性别、以及粪便中所含某种肠寄生虫卵的情况列为调查项目。如要了解与卫生习惯的关系就应增加这方面的项目。

6. 调查表及填表说明：调查表是用来进行调查登记各种原始情况和数字的原始资料。调查表一般有两种格式。

(1) 一览表格式，此种调查表一张可以填写许多个调查单位，如表1。

表1. _____ 县 _____ 公社 _____ 大队 _____ 生产队社员

粪便检查结果记录表

编号	姓名	性别	年龄	职业	检查结果				备注
					蛔虫卵	钩虫卵	鞭虫卵	其他	

_____ 年 _____ 月 _____ 日

填表说明：① ~~年龄~~ 实际年龄（岁）

② 价阶结果为阳性时填“+”号，阴性时填“-”号。

③ “其他”栏填写其他虫卵。

(2) 卡片表格式：此种调查表一张只能填写一个调查单位。此种表格制备比较复杂，但登记的项目可以适当增加，调查内容比较深入。整理也较方便，如表2。

表2

市(县)		区(公社)		居民肿瘤死亡调查卡	
姓名	性别：男、女	实际年龄	岁		
职业	籍贯	住址			
死亡原因	疾病编号	诊断单位			
死亡日期	年 月 日	死亡地点			
	调查人	调查日期	年 月 日		

此外调查表的各项目应当有简单明确的填表说明。

2. 其他：有关调查工作的其他事宜，如组织联系，调查进展，器材准备，表格印制，调查员培训以及交通工具等的安排考虑并列入调查计划内。

(4) 实验设计

实验设计包括实验目的、实验方法、实验对象及其编号、实验指标和实验对象的随机分配方式。例如做动物实验时，须先将动物按相同的体重、性别、雌雄分配成对子（配偶组），然后将每对动物随机分配到实验组和对照组中去。这样可以做到具有齐同一致的对比条件。

常用的随机方法有抽签法。先将配成对子的动物编为1号和2号。如抽出签号的数字为奇数者，则把第1对的1号分到实验组，2号则分到对照组。如抽出签号为偶数者，则把2号分到实验组，1号则分到对照组。

另外，也可以将实验对象完全随机地分成实验和对照两组。就是将实验对象依任何一种无关的标记排列起来，然后采用抽签法，抽奇数和偶数分配到实验组和对照组中去。但此法有时发生分配后两组例数不等，故还需加以调整。

临床工作中，常对患者进行新旧疗法或不同药物的疗效观察，也可应用实验设计的原理来进行。在同一患者身上进行用药前后的对比，统计上称为自身对照。在患者中设立对照组时，可以使用常规有效的药物作为对照，这样可不因观察疗效而延误患者的治疗。

(三) 日常工作记录，登记文件和统计报表

1. 日常工作记录，例如门诊病历，住院病历、体格检查、检查记录。且积累一定数量以后，可用来进行分析一般情况。

如果记录确实，登记完整，这些资料是很有价值的。但医院住院病人资料作统计分析时，须注意是否存在“挑选性”的问题。

2. 登记文件：例如出生、死亡、结婚、职业、肝癌等专用报告卡。由于这些资料收集过程中常有重复、遗漏、含糊不清等问题，故应仔细核实，使资料完整和准确。

3. 统计报表：国家卫生机关制订的卫生报表，由指定的卫生机构，按照一定的程序，自下而上地进行逐级收集。例如全国卫生统计报表中卫生基本情况年报表，医院工作年报表，住院病人疾病分类年报表、居民伤害死亡原因年报表等，都是卫生工作的重要资料。

第二节 统计资料的整理

整理的目的是将收集来的原始资料，进行加工整理使之系统化和条理化，以便于比较和分析。整理的步骤分为：

一、检查核对

目的是检查原始资料中是否有遗漏和错误，以便及早补充和更正。核对的方法可以根据项目之间的关系进行更正，也可以重新调查进行补充。或注意不要随便去掉某个调查单位，以免

产生偏性。

二、设计分组

分组是统计资料整理的基础，为了保证“同质性”和“对比条件一致”的要求，需在进行科学分组。

常用的分组方法有两种，①一种是计数资料，它是按资料的性质来区分的，例如性别、疾病种类、科别、职业、和表示治疗结果的治愈、好转、无变化、死亡等。②另一种是测量资料，它按数量大小来区分的，例如年龄、血压、血红蛋白、血清蛋白、血清钙等。这里着重介绍测量资料分组的方法。

一般测量资料分组，是为了更好地显示出数量变化后的质变，但当分组过粗时，则规律性不能明显突出；反之，如分组过细，又会隐藏内在的特点，故必须合理的进行数量分组。

现以某地居民胸透查出广泛浸润性人般的年龄分组介绍如下。

下：

错误的写法
(岁)

0-10
10-20
20-30
30-40
40-50
50-60
60-70

70及以上

正确的写法
(岁)

0-
10-
20-
30-
40-
50-
60-

70及以上

或写成
(岁)

0-9
10-19
20-29
30-39
40-49
50-59
60-69

70及以上

在年令分组时，~~通常会将各年令情况~~将0—9岁一细内的新生儿、婴儿及幼儿分开，以便于分别观察传染病在青年令组和儿童的不同，故年令也可以不细分，例如：

不满1岁

其中，新生儿	35—
1—	45—
5—	55—
15—	65—
25—	75及以上

以上年令都必须填写实际年令，即满一周岁才算一岁，不满一周岁的算为0岁，满一周岁不到两周岁的称为一岁，余类推。

三、拟查理表

拟制查理表的目的，是将有联系的几个项目包括在一个表内，以便于表达出其相互间的关系，并按照这种关系进行分级记载。

查理表的表式与统计表基本相同，现举例如下：

表3 某种肠寄生虫感染阳性者按性别、年令查理

年令 (岁)	男		女		合 计
	划记	小计	划记	小计	
0—					
10—					
20—					

30 -				
40 -				
50 -				
60 -				
70 -				
80岁及以上				
合计				

四、分组归纳

按照整理表所进行的分组，将调查的原始资料，逐个进行分组归纳计数，常用计数方法有：

(一) 划记法 即划一丁下正字或 | || ||| |||| 来记录各组的次数。这种方法简便，但发生错误以后难以纠正。

(二) 手分卡法 此法用于使用卡片作调查的资料。与划记法基本一致，只是把划记改用手分卡，将卡片依其分组逐个放置在该组内，然后计数各组卡片张数，如发现差错，易于发现和更正。

第二章 相对数

在统计中分析、对比计数资料时，常常在计数、相对数

例如，比较两个生产队社员钩虫感染水平的高低，便不能直接用两个生产队社员中钩虫感染阳性人数来比较，因为两个生产队的社员总人数不同，这时需要将这些数字计算成相对数，即用“感染率”来比较，才能清楚地反映两队钩虫感染的水平。例如第一生产队有社员200人，粪价发现钩虫阳性78人。第二生产队社员320人，粪价发现钩虫阳性85人。若比较两生产队钩虫感染情况，直接用粪价阳性人数绝对数作比较时，则第二生产队85人阳性比第一生产队78人阳性多，而认为第二生产队钩虫感染严重，这就不确切了；因为两个生产队的社员人数不同，第二生产队比第一生产队人数多，如果用“相对数”（感染率）来比较，则：

$$\text{第一生产队社员钩虫感染率} = \frac{78}{200} \times 100\% = 39.0\%$$

$$\text{第二生产队社员钩虫感染率} = \frac{85}{320} \times 100\% = 26.6\%$$

可知第一生产队社员钩虫感染程度比第二生产队严重。另外在比较两种不同疾病的白细胞分类，也同样不能使用各类白细胞的个数来比较，应该计算出各类白细胞所占的百分比来比较。这些百分率和百分比都称为相对数。

一、常用相对数及其计算方法

1. 比

表示事物或现象内各构成部分所占的比重。也称为“构成指标”。习惯上都是以100%作为比例基数。其计算公式为：

$$\text{比} = \frac{\text{事物内部某一构成部分的数值}}{\text{事物各构成部分数值的总和}} \times \text{比例基数}$$

例如调查某地某年居民恶性肿瘤死亡共700例。其中几种主要部位恶性肿瘤死亡例数及其所占比重如表4。

表4. 某年某地居民几种主要恶性肿瘤死亡例数及其构成比

部 位	死 亡 数	构 成 比 (%)
肝	156	22.3
食管	136	19.4
鼻咽	92	13.1
胃	83	11.9
其他	231	33.0
合 计	700	100.0

肝癌死亡占的构成比 = $\frac{156}{700} \times 100\% = 22.3\%$ 。其余类推。

2. 率

表示某种现象在一定条件下实际发生数与可能发生该现象的总数之比。又称为“频率指标”。常以司、千、万、十万分之几表示。

为比例基数。称为百分率、千分率、万分率或十万分率等。其
 计算公式为：

$$\text{率} = \frac{\text{某种现象实际发生数}}{\text{该现象可能发生的总数}} \times \text{比例基数}$$

例如某虫产队共有社员 1780 人，粪便价阶后，钩虫总染
 阳性者有 529 人，其钩虫总染率为：

$$\text{某虫产队社员钩虫总染率} = \frac{529}{1780} \times 100\% = 29.7\%$$

三、对比指标

表示两个有关系的现象，他们数值之比，用以说明两者的
 对比关系。其计算公式为：

$$\text{对比指标} = \frac{\text{甲事物的数值}}{\text{乙事物的数值}} \times 100\%$$

例如：出生性别比例 $\frac{\text{一年内出生男婴数}}{\text{一年内出生女婴数}} \times 100\%$

年增长率百分比 $\frac{\text{某年某事物的总数}}{\text{前一年同事物的总数}} \times 100\%$

第二章 应用相对数时注意事项

一、计算相对数时，分母数值不宜过少，否则计算出的指