

医学统计问答与题解

蒋有年 张志军 编 王广仪 审



-44

人 民 卫 生 出 版 社

医学统计问答与题解

蒋有年 张志军 编

人民卫生出版社出版
(北京市崇文区天坛西里10号)

北京市房山区印刷厂印刷
新华书店北京发行所发行

787×1092毫米 16开本 17印张 396千字

1988年4月第1版 1988年4月第1版第1次印刷

印数: 00,001—6,900

ISBN 7-117-00378-2/R·379 定价: 3.20元

统一书号: 14048·5588

〔科技新书目 158—80〕

编者的话

医学统计学对医药卫生事业的现代化有着独特的作用。它是基础医学、临床医学和预防医学人员在从事教学、科研和实际工作中不可缺少的工具。它既可以用于对数量的分析,帮助我们制订医疗卫生工作计划,总结经验,为检查、考核工作提供依据;又可以辅佐我们进行医学科学研究和疾病防治的实践,以达到用较少的人力、物力和时间,获得较大的成效,从而有利于为医学科学积累更有价值的资料。在进行研究设计、撰写医学论文时,能否恰当地运用医学统计方法,也是左右其质量的一个关键。因此,为了做好本职工作、更有成效地从事医疗卫生科研和实际工作或有分析地应用他人的科研成果等,就需要学习并能掌握运用医学统计学的基本知识和技能。

为便于广大基层初、中级医疗卫生人员在工作实践中进一步学习、掌握和运用医学统计方法,以及提供自学者和医学院校学生的参考资料,也考虑到晋升考核的需要,特编写本问答与题解。

考虑到读者学习和复习的方便,本书在编排上仍遵循医学统计学教科书的顺序。在选题上注意结合实际和力求有代表性,内容立足通俗,解题方法步骤强调层次。每章大体都包括三个部分:①基本概念、基本知识问答;②基础应用题题解;③汇总性表解。在书末还附有部分练习题(有答案),尽可能涉及到不同专业的需要。

在编写过程中主要参阅了目前国内的医学统计教材、医学期刊,也把我们在教学和工作实践中的一些体会纳入进去。我们在编写中得到单位领导和同志们的大力支持,也受到人民卫生出版社编辑同志的热情帮助和指导,白求恩医科大学王广仪教授审阅了全书,在此一并致谢。希望本书能为普及和推动医学统计知识,为我国医疗卫生事业的发展做出微薄的贡献。但是由于我们水平所限,缺点错误在所难免,切望广大读者批评指正。

编者

目 录

第一章 医学统计学的基本概念	1
第一节 医学统计学的概念、对象和原则 (问答 1~4 题)	1
第二节 统计学的几个基本词汇解释 (问答 5~9 题)	2
第二章 统计资料的类型、来源和整理 (问答 10~14 题)	5
第三章 平均数与标准差	11
第一节 算术平均数和几何平均数 (问答 15~17 题)	11
第二节 均数、几何均数、中位数题解 (18~24 题)	14
第三节 百分位数及其应用题解 (25~28 题)	19
第四节 标准差及其应用题解 (29~35 题)	22
第五节 变异系数及其计算题解 (36~40 题)	25
第四章 正态分布	29
第一节 正态分布的概念 (问答 41~43 题)	29
第二节 正态分布题解 (44~47 题)	31
第五章 均数的抽样误差与均数差异的显著性检验	33
第一节 基本概念 (问答 48~53 题)	33
第二节 均数差异显著性检验的基本概念和一般步骤、注意事项 (问答 54~57 题)	36
第三节 两均数差别的显著性检验题解 (58~66 题)	41
第四节 两种检验和两类错误 (问答与题解 67~69 题)	51
第六章 多个样本均数差异的显著性检验——方差分析	54
第一节 基本概念 (问答 70~75 题)	54
第二节 方差分析题解 (76~82 题)	60
第七章 相关与回归	72
第一节 相关 (问答与题解 83~90 题)	72
第二节 回归 (问答与题解 91~96 题)	79
第三节 等级相关 (秩相关) (问答与题解 97~99 题)	84
第四节 多元分析 (问答与题解 100~102 题)	86
第八章 正常值范围的确定 (问答与题解 103~115 题)	94
第九章 相对数与率	108
第一节 相对数、率及标准化率 (问答与题解 116~131 题)	108
第二节 二项分布、泊松分布及率的抽样误差 (问答与题解 132~144 题)	119
第十章 χ^2 检验 (问答与题解 145~154 题)	127
第十一章 非参数统计方法 (问答与题解 155~167 题)	135
第十二章 cpd 分析 (问答与题解 168~175 题)	146
第十三章 序贯检验	153
第一节 基本概念 (问答 176~178 题)	153
第二节 开放型序贯检验题解 (179~184 题)	157
第十四章 统计表与统计图	164

第一节 基本概念 (问答 185~187 题)	164
第二节 常见统计图、表的种类和绘制要点 (问答 188 题)	167
第三节 统计图、表错误的修正及应用题解 (189~200 题)	167
第十五章 医学研究设计基本知识 (问答 201~232 题)	180
附表	208
1 标准正态曲线下的面积表 (单侧面积)	208
2 t 值表	209
3 F 值表 (方差齐性检验用, 双侧)	210
4(1) F 值表 (方差分析用, 单侧)	211
4(2) F 值表 (方差分析用, 单侧)	212
4(3) F 值表 (方差分析用, 单侧)	213
4(4) F 值表 (方差分析用, 单侧)	214
5 q 值表 (单侧)	215
6(1) 相关系数界值表	216
6(2) 相关系数界值表	217
7 等级相关系数界值表	218
8 正态性 D 检验界值表	219
9 计数资料单向序贯检验边界系数表	220
10 计量资料单向序贯检验 $\frac{s_1}{\sigma}$ 值表	222
11 配对计数资料序贯检验边界系数表	223
12 配对计量资料序贯检验边界系数表	223
13 两样本率比较时所需样本含量 (单侧)	224
14 两样本率比较时所需样本含量 (双侧)	225
15 配对比较 (t 检验) 时所需样本含量	226
16 两样本均数比较 (t 检验) 时所需样本含量	227
17(1) 随机数字表	228
17(2) 随机数字表	229
18(1) 百分率的可信限 ($1 \leq n \leq 50$)	230
18(2) 百分率的可信限 ($1 \leq n \leq 50$)	231
18(3) 百分率的可信限 ($1 \leq n \leq 50$)	232
18(4) 百分率的可信限 ($50 \leq n \leq 100$)	233
18(5) 百分率的可信限 ($100 \leq n \leq 1000$)	234
19 泊松分布均数的可信限	235
20 χ^2 值表	236
21(1) $L_4(2^3)$ 正交表	237
21(2) $L_8(2^7)$ 正交表及交互作用表、表头设计表	237
21(3) $L_{16}(2^{15})$ 正交表及交互作用表、表头设计表	238
21(4) $L_{12}(2^{11})$ 正交表	239
21(5) $L_9(3^4)$ 正交表	239
21(6) $L_{27}(3^{13})$ 正交表及交互作用表、表头设计表	241
21(7) $L_{16}(4^5)$ 正交表	242
21(8) $L_{25}(5^6)$ 正交表	242

21(9) $L_8(4^1 \times 2^4)$ 正交表	243
21(10) $L_{12}(2^3 \times 3^1)$ 正交表	243
21(11) $L_{16}(4^4 \times 2^3)$ 正交表	244
21(12) $L_{16}(4^3 \times 2^6)$ 正交表	244
21(13) $L_{16}(4^2 \times 2^9)$ 正交表	245
21(14) $L_{16}(4^1 \times 2^{12})$ 正交表及表头设计	245
21(15) $L_{16}(8^1 \times 2^8)$ 正交表	246
21(16) $L_{18}(2^1 \times 3^7)$ 正交表	247
22 相关顺序检验的K 界值	247
医学统计练习题(附答案)	248

第一章 医学统计学的基本概念

第一节 医学统计学的概念、对象和原则

(问答1~4题)

1 题 什么是医学统计学?

解 医学统计学是数理统计在医学中的应用。它是根据数理统计的原理和方法,来分析和解释医学中各种现象和数量资料的科学。它是在医学发展过程中逐渐与数学结合而形成的,并且仍在不断发展。

在医疗预防工作中,正确地运用统计分析方法,有助于我们认识事物的规律性。为了有的放矢地开展工作,有效地提高工作质量就必须掌握基本情况,做到心中有数,就要借助于医学统计学的帮助。在医学科研、疾病防治工作中,无论是拟订计划、总结工作、撰写论文或工作成绩的正确评价,以及找准差距、明确方向等方面,都必须有计划地收集资料 and 进行合理的统计分析。因此,医学统计方法是医学工作者一定要学习和掌握的有用工具。

2 题 医学统计学的研究对象和主要内容是什么?

解 有变异的事物是统计研究的对象。医学的研究对象普遍存在着变异,这种变异甚至是较大的。如同患一种疾病,不同的患者病情轻重、病程长短、疗效好坏会有所不同;就是同一病人,其症状和体征在不同时间、不同条件下亦会有变化,同年龄同性的健康人,其解剖、生理、生化指标,如身长、血压、白细胞数、尿糖含量等也不相同,医学统计学就是研究这些变异的规律和特征的。没有变异的事物不是统计学研究的对象。

医学统计的基本内容主要包括两大部分:实验设计与数据分析。

实验设计是指在试验研究工作未进行之前,应用医学统计原理和丰富的专业知识,对所需要研究的问题进行的设想与计划安排。亦即要制订合理的试验方案,正确地选择试验对象,正确地进行分组和对比,以用较少的人力、物力和时间,最大限度地获得有效而可靠的信息,得出科学的结论。实验设计的主要内容是明确研究目的、确定研究方法和步骤。一项缺乏周密和完善的实验设计,不仅可能对工作造成不应有的浪费,而且足以减损研究结果的价值。因此,搞好实验设计是实验和调查研究工作的重要环节。在后面将有专章讨论。

数据分析亦即统计描述与统计推断,或称统计学处理。这是医学统计学的另一项重要内容。它是对通过实验或调查收集到的数据进行科学的分类、整理、分析与处理;用一定的统计方法对全部同类性质的研究对象进行科学的逻辑推理;研究这些数据资料的表达,计算相应的统计指标与绘制统计图表;并研究抽样误差估计、数据可靠性的评价,进行差异的显著性检验,以及研究数据之间的关系等方面的内容。

3 题 什么是医学统计学的基本原则?

解 医学统计学的基本原则有以下几点:

首先,坚持辩证唯物主义的原则。要用辩证唯物主义的观点和方法,严肃认真、实

事求是的态度看问题，切忌主观臆断、凭表面现象和数据分析问题和做结论。

第二，随机化的原则。为了使样本对总体有较好的代表性，防止主观随意偏差，并使其抽样误差的大小能用统计方法来估计，在抽样研究时就必须遵守随机化的原则。在实施观测对象分组或安排实验顺序时，亦应用随机方法。

第三，观察数量足够的原则。既然统计的对象是有变异的事物，那么就不能孤立地研究个别事物，而要在一定数量观察的基础上进行，以充分显示出差异的本质，才便于从诸现象里研究事物间的相互关系，阐明事物客观存在的规律。在医学研究工作中，往往是通过某事物的一部分（样本）去估计其所在的全体（总体）的特征，目的是由样本推断总体，从特殊到一般，对所研究的总体作出合乎逻辑的推论，得到对客观事物本质的和规律性的认识。医学统计学在某种意义上讲，就是研究医学中以样本推断总体的一门学问，以便对同类事物加以估计和预测，借以指导医学实践，推动工作的进展。

第四，同质性、可比性的原则。统计研究还必须注意被研究事物在性质上的同质性、可比性。如在研究某种新药对菌痢的疗效时，所有研究对象都必须是确诊为菌痢的病人，不包括疑似或根本不是菌痢的人在内。务使研究对象是同质的，否则研究结果就不能反映这种新药对菌痢的疗效。可比性是指同质的某项指标是在相一致的时间、地点、条件下获得的，才能做比较。如比较两个乡之间的流感发病率，要在季节、地区、卫生、经济条件相类似的情况下，才能相互比较。如这些条件不同，就缺乏可比性。

第五，不能只凭常识或数字的表面值，不经必要的统计处理就武断地下结论。否则，会导致错误。要防止“统计无用”和“统计万能”的观点，就是说既不能认为没有运用统计学知识也照样可以搞医学研究，又不能以为只要运用了统计学方法就会解决一切问题。统计方法是一种处理数量资料的工具，它决不能改变事物的本来面目，而把不存在的规律创造出来，它只能起到揭示、反映规律的作用。不遵循统计学原则所收集到的资料，即使经过统计学处理，也将是得不出可信的结论，甚至以讹传讹贻误工作，而统计学弥补不了这种弊病。凭主观愿望或滥用统计方法去凑合预定的结论，是错误的。

4 题 怎样才能学好医学统计学？

解 这方面并无成熟的方法，只谈谈我们的粗浅体会。一是，要有一定的数学知识。一般来说要具有中学的数学水平。因为在我们进行实验设计，收集、整理、分析资料时，离不开数的运算、处理，而且还要通过一定的数量去表达事物的本质，这当然要有一些数学知识才行。尽管广泛使用了计算器和计算机，但也离不开数学基本知识。二是，一般来讲，对于初学者、从事基层卫生医药工作人员在学习医学统计方法时，没有必要去推导统计公式的来源和深究其数学原理，要注重理解统计学的基本概念，确实弄清基本原理，掌握收集、整理与分析资料的基本方法。要搞清各种统计方法、公式的适用条件和正确用法。三是，要重视边学边用，在实践中学。在学习中学切忌照葫芦画瓢、蜻蜓点水、一曝十寒以及脱离专业实践。只要坚持下去，逐步地掌握和运用它，不断提高用统计方法去解决和分析问题的能力，当能推动工作的发展，受益非浅。

第二节 统计学的几个基本词汇解释

（问答 5~9 题）

5 题 什么叫总体、样本和观察单位？

解 总体是根据研究目的确定的研究对象的全体，亦即结论所要包括的范围、同质个体所构成的集体。总体所包括的个体数可以是无穷大，而且总体往往只是设想的或是抽象的概念。不同的研究目的决定了各自对应的总体。例如，我们欲观察某药对溃疡病的疗效，那么所有的溃疡病病人就是这项研究的总体，他们都是同质的。其中每个溃疡病人叫作个体，即总体中的每个同质的“单位”。在研究过程中，常常不可能，也不需要去观测总体中的每个个体，而只能对一部分个体进行观测。从同一总体中随机抽取部分个体的过程称为抽样。所抽得的部分个体就称为样本。样本所包含的个体数目便是样本含量。样本含量的大小，要根据指定的方法来确定。

观察单位是所研究的同质总体中的每一个个体。它可以是一例患者、一个居民、一个家庭、一个行政或生产单位，也可以是“人次”或采样点、一个被试动物、一个血液标本、一份水样等。

在医学科研中，我们所期望知道的是总体的特征，而不是样本的特征。可是在具体实践中，我们手头所掌握的资料经常是一个样本。例如，用某中药治疗慢性支气管炎病人 200 例，有效率为 80%。此时观察的只是含量为 200 例的一个样本。而观察的目的却是该中药对所有这种病人的疗效，亦即如果再用这种中药治疗其他慢性支气管炎病人，是否也会获得同样为 80% 的疗效呢？这就需要以样本的疗效去估测总体的疗效。而如何正确地用样本去推测总体，这是统计学所要解决的主要问题之一。

6 题 什么叫抽样误差、抽样研究？抽样研究的目的是什么？

解 由于总体中的各个个体存在着变异，因此在同一总体中随机抽取若干个个体数相同的样本，各样本均数（或各样本率）不仅会跟总体均数（或率）不尽相同，而且各样本均数（或率）间也会有所不同。这种由于抽样而引起的差异，在统计学上称为抽样误差。抽样误差是抽样研究中必然存在的，但可以把它压缩到最低限度。抽样误差是有一定规律的，研究和运用抽样误差的规律以指导研究设计与资料分析，是医学统计学的另一重要内容。

在总体中随机抽取样本进行研究，用样本指标估计总体指标的方法称为抽样研究方法。抽样研究的目的在于用样本去推论总体。正确的运用统计处理方法，就可以识别在实验研究中哪些差别是有实际意义的，哪些差别可能仅是由于偶然性所造成的，从而有助于得出一个正确的结论。

7 题 什么是随机事件和随机化？

解 因为统计研究的对象存在着变异，而变异的出现是由于许多内外因素偶然性综合所致，所以统计研究的各种现象的表现结果是一种随机事件。随机事件（或偶然事件）是指一次试验不能确定，而在一定数量重复条件下才呈现出统计规律性的事件，简称事件。亦即在一定条件下可能出现也可能不出现的现象。个别随机事件的出现带有偶然性，好象无规律可循。但是，若对大量的同类随机事件进行观察与试验，我们就会发现随机事件的发生具有必然的规律性。例如，掷硬币，若掷一次时则可能是得币值的一面，也可能是得国徽的一面。但是，若大量重复投掷硬币，则会发现得币值一面的次数约为投掷总次数的一半，即可能性为 50%。概率论与数理统计就是从数量上研究大量同类随机事件规律性的科学。

随机化，就是在抽取样本以前要使总体中每个单位都有同等的被抽取的机会，进而

使样本对总体有较好的代表性，并使其抽样误差的大小可以用统计方法加以估计。在实验与调查研究时，要将受试对象分配成几个组，这时也必须用随机方法使每个对象都有同等的机会被分配到各组去，这样就不至于人为地造成各组间对象的不齐同。抽样研究和抽样分配都遵守随机化的原则，所获得的资料才适合统计处理的需要。这是因为一般数理统计的计算方法是在随机抽样的基础上推演而来的。

8 题 什么是概率？

解 概率是指某事件出现的可能性大小的量，亦称为机率、或然率。因为，对于随机事件，仅指出其发生的偶然性是不够的，重要的是指出事件发生的可能性大小。在一般情况下，某事件发生的概率是不可能准确得到的，常以当试验次数 n 充分大时，某事件出现的频率作为该事件概率的近似值。如投掷几枚硬币，得国徽一面是随机的，结果可能是 45% 或 60%。这是因为投币数太少，偶然性所致。但是，随着投掷硬币个数的增多，得国徽一面的频率越来越呈现出稳定性，即得国徽一面的频率越来越稳定地接近于定值 50%（即 0.5），只是偶尔发生较大的偏差。这个稳定的定值便是硬币得国徽面的概率。医学研究中某事件的概率也同此理。

概率的大小，亦即某事件出现的可能性的的大小，在数学上是用分数、小数或百分数表示的。概率的符号是“ P ”。概率的数值波动在 0 与 1 之间。某一事件必然不发生（不可能事件），则该事件发生的概率为 0；某一事件必然发生（必然事件），则该事件发生的概率为 1；某事件发生的概率愈近于 0，表示该事件发生的可能性愈小，而 P 愈接近于 1 则说明该事件发生的可能性愈大。随机事件的概率表示了事件的客观规律性，它反映事件在一次试验中出现的可能性的的大小。若某事件的概率很小，如 $P < 0.05$ （即 5%）或 $P < 0.01$ （即 1%），表示某事件在一次试验中出现的可能性很小或者极小，以至实际上可以认为是不可能事件。这种将小概率事件在一次试验中看成是实际不可能出现的事件称为“小概率事件实际不可能性原理”。这个原理是统计方法上进行差别显著性检验的基本原理，统计分析的许多结论都是建立在概率大小的基础上的。

9 题 什么是统计量和参数？

解 由样本观察值所计算出的统计指标称为统计量。例如为了了解正常成年男子身高而进行身高测量，由所测得的数据计算出身高的平均数。这是样本平均数，是一个统计量。样本标准差、样本率等也是统计量。统计量的符号用拉丁字母表示，如 $\bar{X}$ 表示样本平均数， s 表示样本标准差， P 表示样本率。

总体平均数、总体标准差、总体率等总体的统计数值称为参数。参数往往是由样本的统计量估计得到的。参数的符号用希腊字母表示，如 μ 表示总体平均数， σ 表示总体标准差， π 表示总体率，等。

（蒋有年）

第二章 统计资料的类型、来源和整理

(问答 10~14 题)

10 题 试述统计资料的类型和统计方法(表解)。

解 见表 2-1。

表 2-1 统计资料的常见类型和统计方法

类 型	概 念	常用的统计方法
计量资料	是对每个观察单位用定量法测定其某项指标所得数值大小的资料。 一般用度量衡等单位表示,如胸围(cm)、体重(kg)、浓度(mg/L)、血压(mmHg)、红细胞数(个/mm³)。	1. 指标法: 平均数、标准差、变异系数、正常值范围的计算或确定等。 2. 估计法⁽¹⁾: 总体均数的估计、频数分布的估计等。 3. 比较法⁽²⁾: 均数差异的比较、两组几何均数的比较、秩和检验等。 4. 相关法⁽³⁾: 相关及回归分析。
计数资料	是先将观察单位按性质或类别进行分组,再清点各组观察单位的个数所得的资料。 如对某个乡的全体农民进行沙眼检查,然后将他们分成患沙眼组、未患沙眼组,再清点各组的相应人数。又如调查人群的血型分布时,先按照A、B、AB、O 四型分组,再清点各血型组的人数。	1. 指标法: 率、构成比、相对比等。 2. 估计法: 总体率的估计、总体标准化率的估计、动态趋势等。 3. 比较法: 率的差异比较、两组标准化率的比较、秩和检验等。 4. 相关法: 两种方法的结果是否一致、两种方法效果能否分出优劣、两种方法结果有无关联、配对调查中是否存在因果关系,以及等级相关等。
等级资料 ⁽⁴⁾	是既具有计数资料的特性,又兼有半定量的性质的资料。是将观察单位按某项指标的等级顺序分组(具有半定量性质),然后再清点各组观察单位的个数所得的资料。 如血压按测定值的不同,分为低血压、正常血压、临界高血压、高血压组,再清点各组的人数。其它如疗效等级、血清稀释度分组资料等。	如cpd分析法和秩和检验等。 注意: 1. X² 检验(独立性检验、关联性检验),适用于无序计数资料(如血型、性别等无自然顺序)、2×c 表资料,而对有序计数资料(如疗效等级分组资料、列联表资料)不适用。 2. cpd分析法(交叉积差法),尤适用于检验有序计数资料(等级分组资料、R×c 列联表资料)。等级分组指划分 3 组及以上的资料。

注(1) 估计法: 指对单个率(构成比)或一组率的估计。单个率是某单一现象的出现率,一般用可信限估计其相应总体的率之可能范围。对于一组动态数列进行分析,实质上就是对单个率的分析在时间上的延伸;对于计量资料,运用估计法是由单个样本平均数,以可信限估计其相应总体平均数的所在范围。估计法的特点是要求资料本身具有独立性,分析的结果只限于说明某一个率、某一组率或某一个平均数、某一组平均数。该法的特点是在于“估范围”。

(2) 比较法: 对于计数资料是用于两个率或多个率间差异的比较;对于计量资料是用于两个平均数或多个平均数间差异的比较,或一组差值、一组差异等级的比较。比较法的特点是: 分析目的在于“比大小”。

(3) 相关法: 是指分析两组结果之间的关系。特点是在同一样本的范围进行两个指标之间的关系的分析,或两样本数量上的相关关系。

(4) 等级资料: 根据分析研究目的的需要,可以把计量资料变换为计数资料或等级分组资料。例如,白细胞总数是属于计量资料,但可以按白细胞总数正常(4000~10000个/mm³)与白细胞数不正常(多于10000或少于4000个/mm³)分为两组,再清点各组人数,就成为计数资料。如果是按白细胞总数过高(多于10000个/mm³)、正常(4000~10000个/mm³)、减少(少于4000个/mm³)分为三组,再清点各组人数,这样就形成按等级分组的资料了。

11 题 试述统计资料的来源有几? 有什么要求?

解 统计资料的来源主要有二:

(一) 经常性资料

1. 工作记录

(1) 日常医疗卫生工作的原始记录。如门诊、住院、接生、健康检查、传染病处理等记录。

(2) 专用报告单。如出生、死亡、传染病、职业病、地方病、肿瘤、外伤等报告单或报告表。

以上资料均是医疗卫生机构例行工作的凭证和工具，也是业务管理和科学研究的重要资料。这些资料可以用来作为评价医疗预防和卫生防疫工作的质量和效果，分析疾病种类和病伤死因的数量、比例、变化，以及探讨发病规律的根据等。因此，这些记录和报告，必须认真填写和注意积累、保存，并予以充分利用。

2. 统计报表。诸如反映卫生工作基本情况的年报表，疫情旬、月、季、年报表，职工病伤死因年报表，计划生育年报表，疾病发病及休工情况报表等。

统计报表是定期取得系统的、全面的、准确的统计资料的主要形式。这是了解居民健康状况，拟定医疗卫生工作计划与措施，检查、总结工作和进行科学研究的基本资料和依据。

(二) 一时性资料

这是针对某些专门问题的资料。如了解详细的死因分类，分析致病原因，研究生长发育，调查环境卫生、职业病，观察某种实验研究结果等。在进行这些调查时，单靠经常性资料是远远不够的，而必须临时组织人力、物力进行专题的调查或实验研究，以获得所需要的资料。

一时性资料按其收集的方式和特点，又可分为现场调查资料和实验研究资料两类。前者是对现场实际已经发生和存在着的情况进行调查，如某幼儿园儿童蛔虫感染率的调查。后者是在实验条件下对某种预期效果的研究，如某新药对某病的疗效观察。经常性和一时性资料两者既有区别，又是相辅相成的。

对原始统计资料主要有三点要求：

(1) 完整：是指应该填写的项目都要填写上，避免重复或遗漏。

(2) 准确：是指原始资料的收集和登记应严格按照要求，实事求是地填写，不得虚假、隐瞒或不填。

(3) 及时：按设计方案的要求，及时填写才可保证在某一特定的时间、地点条件下的具体情况。

原始资料是统计的“原料”，只有保证以上三点要求，才能真实反映情况，确保资料的科学性和可靠性。

对于医院的病史资料，要注意几方面的问题：一是，大医院往往收住病情较重的病人和疑难病人较多，用他们代表这种疾病病情的普遍性是欠妥的。因为这属于有“挑选性”的样本，它只能代表具有同样病情的总体。在进行不同医院的资料对比时要慎重；二是，在不同时期疾病治愈率会不同，如某医院由于医院病床数增多，收住轻患较多，以及由于各种条件的改善、医护技术业务水平的提高，对疾病的诊断水平和治愈率也会有所提高。所以在比较不同时期的资料时，要注意这种情况；三是，医院的病史资料不能包括所有病例，如许多病例不需或没能住院治疗，因此医院的病史资料并不全面。一个地区医院的病史资料总和，不会等于同时期该地区的总发病人数。所以只根据门诊或

住院病例去估计居民的发病数、发病率和疾病的严重程度是不行的。

12 题 统计资料的整理包括哪些基本步骤？其主要内容是什么？

解 为了了解事物总的特征和发展情况，使资料系统化、便于进一步统计分析而揭示被研究事物的规律性，必须对收集到的大量且分散的原始资料进行科学的分组归纳，这个过程称为统计资料的整理。

收集的资料在没有整理分析前叫原始资料。对原始资料的整理工作可分为检查资料、设计分组、拟整理表和归纳四个步骤，分述如下。

（一）检查资料

首先要检查资料的完整性，逐一查看原始记录有无遗漏、重复；其次是检查资料的正确性，主要检查各个项目的填写是否正确，各项目之间有无矛盾和数字不合理之处；最后，对查出的问题应当作合理的修正、补充或删除，必要时进行复查。在专题调查时，应随时核对资料。

（二）设计分组

在检查完资料之后，便可进行分组整理。分组是根据各项目的特性，将同质（性质相同）的资料归纳到一组，以使资料系统化。分组时，一定坚持“同质”的原则，才能显示资料内部的规律性。分组方法有两种：

1. 质量分组。按事物的性质或类型来分组，如疾病分类、死因分类；人口按性别、职业按工种等分组；疗效按治愈、好转、死亡等分组；粪检虫卵按阳性、阴性分组等。不同质的资料不能并在一组，如研究心脏病与年龄有无关系时，不能将风湿性心脏病与高血压性心脏病并在一组，因为两者在年龄分布上不同，合组就会掩盖这一规律性。

2. 数量分组。在质量分组的基础上，再按数量的大小来分组。如年龄、血压、脉搏、身高、红细胞数、白细胞数等。数量分组时，对分组的要求是：

① 组数要适宜。组数的多少和分组的粗细，取决于资料的性质和资料的多少，以能说明资料的规律性为准。分组太粗、组数太少，常可隐蔽资料的特点，如在研究一般发病率时，若以 0~19 岁划为一组就看不出婴、幼儿和青少年的发病特点，常从 0 岁开始，每 5 岁为一组。反之，分组太细、组数过多，会使每组例数过少，也将反映不出资料的规律性。一般要分为 8~15 组。常按等距分组，但为了反映各年龄组的发病特点亦可采取不等距分组，如急性传染病在儿童中多见，故 10 岁以内可按岁分组，10 岁后按每 5 岁分组等。

② 分组的界限要清楚。组界不清时会造成误解，归组资料时会发生重复或遗漏。如在职业病调查工作中，按工人的工龄进行分组：0~5、5~10、10~15……（年），就存在着组限不清的问题。若某工人的工龄是 5 年，应归放在 0~5（年）组还是 5~10（年）组中？这便会发生重复或遗漏。正确的分组界限应是：0~、5~、10~、15~……（年），或 0~4、5~9、10~14、……（年）。

（三）拟整理表

整理表是过渡性表格，是用于原始资料整理归组的表格。根据研究的目的和要求，应有联系的几个项目制订在一个表内，使相互关系和事物的规律性表达出来。例如某特区普查农民结核病，在分析结核病与年龄和性别的关系时，就可以拟订如表 2-2 的整理表。

表 2-2 某特区农民结核病普查结果整理表

年龄组(岁)	男		女		计	
	检查人数	病例数	检查人数	病例数	检查人数	病例数
0~						
5~						
10~						
⋮						
70岁及以上						
合计						

在编制项目时,应按次序合理排列,如时间按先后、数字按大小、疾病按分类。在设计分组标志时,应尽可能细致清楚,要包括资料可能出现的各种情况。最后还要有合计、总计等项目,以便于互相核对。

(四) 归纳汇总

归纳汇总是按整理表的分组,将原始资料逐个分别归入各组,再按组合计频数。一般采用划记法或分卡法。

1. 划记法。在将资料过录归组时,用划“一、丁、下、正、正”或“Ⅰ、Ⅱ、Ⅲ、Ⅳ、Ⅴ”来记录,之后计数各组的频数。此法简便易行,但易出错,出错后只好返工重划。因此,划记时要小心谨慎,至少要划两遍以资核对。常用于资料数量不多,项目较少时的归组。

2. 分卡法。此法用于使用卡片作调查的原始资料。由于每个观察单位登记在一个卡片上,所以只要将登记后的卡片分别归入各组,再清点每组卡片的张数,就是该组的观察单位频数。此法易于核对和检查错误,故常用于资料数量较多、项目较多时的归组。

当调查表为“一览表”时(就是多个观察单位登记在同一张原始调查表上),须先将各观察单位的原始记录分别转抄到“过录卡”上,然后再用分卡法归组。过录卡的式样如表 2-3。表上的编号与一览表上观察单位的编号同。卡内各格的号数固定代表一个分析项目。如 1 号格代表性别,2 号格代表年龄等;也可代表检验结果,如“-”、“+”(阴性、阳性)。

资料整理好后,即可进行统计指标的计算和分析,列出统计图表,最后写出报告。

表 2-3 过录卡片式样

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20

13 题 试述统计工作的基本步骤及主要内容。

解 统计工作的基本步骤是资料的收集、资料的整理、选用统计指标、进行统计分析和具体应用五个步骤。各步骤的主要内容简述如下：

第一步 资料的收集。包括对收集资料的设计和实施。原始资料应是准确的、完整的、及时的，资料的来源、种类和要求（见 10、11 题）。

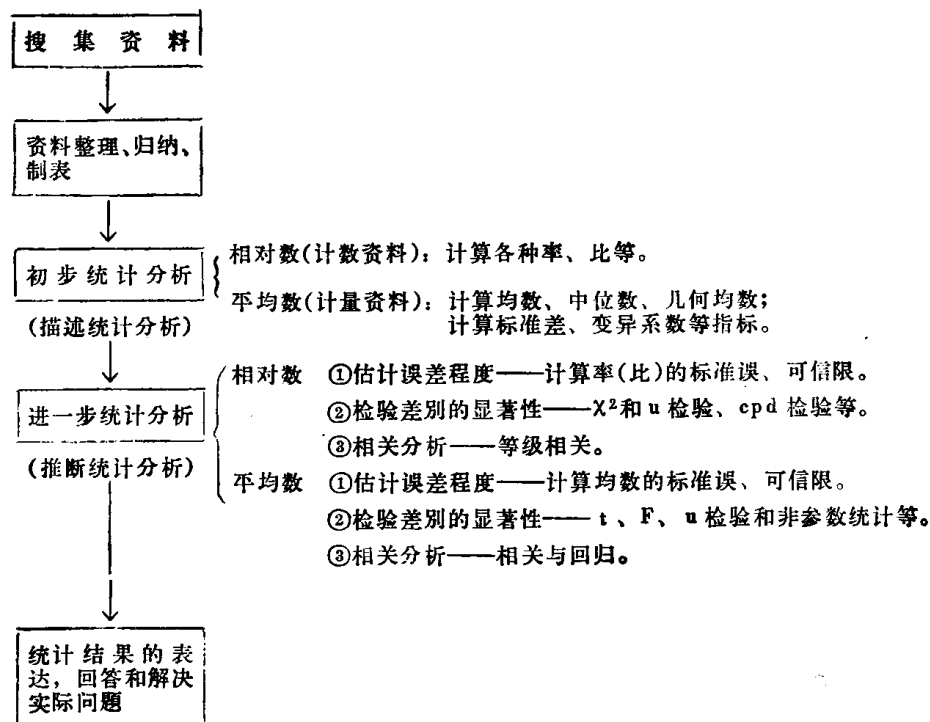
第二步 资料的整理。因为原始资料的数据一般是分散的、零星的，只表明被观察对象各自的特征，需要通过整理资料才能找出它们的共同点，以显示全部被观察对象总括的特征（见 12 题）。整理资料所用的统计表的设计和统计图的绘制方法，见第 14 章。

第三步 选用统计指标。统计指标的效用是以一个或几个简单的数字集中地、明确地反映资料的特征。常用的统计指标见本题后部。

第四步 进行统计分析。统计指标不是孤立的，往往要从它们的分布、动态、对比、相关等方面进行分析，去粗取精、去伪存真、由此及彼、由表及里，才能从错综复杂的数据中，提炼出符合事实真相的结论，找出规律性来。为此，要对资料进行统计指标的初步分析，进一步再作抽样误差的分析，对计数资料统计指标（如率）、计量资料统计指标（如平均数）作差异的显著性检验（如 χ^2 检验， t 检验，或 u 检验等）。需要时，还要对计量数据进行相关、回归等分析。这些统计分析，将在有关的题解中例述。

第五步 具体应用。一般统计工作都是依据一定的业务需要而进行的，这样，统计的结果就要密切联系实际，不然则是以空对空的数字游戏。也就是要用统计的结果去说明专业上的问题，为医疗卫生工作服务。我们将在大量实例中介绍怎样将应用统计分析得到的结果与专业实际相联系。

现将统计处理过程概括示意如下，



14 题 怎样表达统计结果(表解)?

解 见表 2-4。

表 2-4 常用统计结果的表达形式

统计结果类型	表 达 形 式	举 例
相对数 (计数资料)	列出组别、被检查人数、阳性人数、阳性率；疾病分型(或组别)、治疗人数、治愈人数、治愈率；在阳性率下面附以卡方(χ^2)值或cpd法的统计量值、P值,或注明各组相差不显著、显著或非常显著。在文字说明中可进一步指出某组或某几组的阳性率(或某种率)最高,哪几组其次、最低;关于 χ^2 检验或其他检验的演算过程及所用公式不必列出(其他见表2-1“注意”)。	某地检查成年男子125,000人,高血压9,625人,高血压检出率7.7%。 又如,用甲、乙两种药对同种疾病各治100例,治愈率分别为92%、65%。 $\chi^2=21.6$, $P<0.01$ 。则甲、乙两药疗效差异有极显著性意义,可以认为甲药优于乙药的疗效。
计量资料	小样本($n<30$)中,必要时可列出各组原始数据,附以平均数、标准差。有时可用“平均数±标准差”的形式,在文章中注意不要与“平均数±标准误”相混。 大样本统计结果,可列出频数表和合计总人数,并附以平均数与标准差。	如某年某地测得300名10岁男孩身高的结果,平均身高为126.1cm,标准差为5.32cm,或 $126.1\pm5.32\text{cm}$ 。
比较两均数的差别	列出均数及其差别、t值、P值,并可注明两均数相差为不显著、显著、或非常显著。演算过程不必列出。并一定写出观察例数。	如测得两组(各10例)对象的血磷均数 $\bar{X}_1=4.7$, $\bar{X}_2=3.4\text{mg}\%$, 差别为1.3mg%, $t=2.5$, $P<0.05$ 。此差别有显著性意义。
比较三个或三个以上均数的差别	列出各组例数、均数、F值、P值,可把方差分析表列出,并可注明相差不显著、显著或极(非常)显著。	如用三种方法各治某病5例,其平均治愈日数各为5、4、9天, $F=5.38$, $P<0.05$ 三种方法治愈日数有显著不同。
分析两变量间的关系	小样本可列出原始数据,大样本可列出相关表。列出直线回归方程及t值、P值或加注其显著性情况。若t检验不显著就不列出回归方程,只说明“经直线回归处理结果,二者无显著关系存在”;求相关系数时,列出r值、t值、P值,或加注其显著性情况。若r不显著,则只写“经计算相关系数后,二者无显著关系存在”。	如测得某克山病区10名健康儿童发硒与血硒值(暂略),相关系数 $r=0.872$, $P=0.0005$,认为发硒与血硒含量间有相关关系存在。
正常值的确定	说明确定正常值所用方法,并列单侧(上限或下限)正常值百分界限的数值;或列出双侧正常值百分界限的范围。均应注明正常值的单位。	如用百分位数法测得某市健康成人血铅单侧正常值95%上限为 $39\mu\text{g}/100\text{g}$ (双硫腙比色分光比色法)。

(蒋有年)

第三章 平均数与标准差

第一节 算术平均数和几何平均数

(问答 15~17 题)

15 题 什么是平均数、算术平均数、中位数和几何平均数?

解 平均数是分析计量资料常用的一种统计指标。平均数是算术平均数、几何平均数、百分位数、中位数、调和均数和众数的通称。平均数是用以说明一组同质观察值的平均水平或集中趋势。如我们在研究儿童身体发育状况时, 尽管是同年龄、同性别、同民族, 但每个人的身高(或体重)并不完全一样, 为表达这样一批“同质”人的平均身高(或体重), 就可通过计算他们的平均身高(或体重)值, 来反映平均身高水平。在医疗卫生工作中最常用的平均数有算术均数(简称均数)、几何均数和中位数。

均数是各观察值的平均值, 符号为 $\bar{X}$ 。

几何均数是 n 个观察值的乘积开 n 次方所得到的根。符号为 G (或 $\bar{X}_G$)。

中位数是一组观察值按由小到大的顺序排列后, 位次居中的那个数。它的上下(或前后)各有相等的频数观察值分布着。符号为 M 或 m_d 。

16 题 试述常用平均数的统计方法、应用及注意事项(表解)?

解 见表 3-1。

17 题 怎样正确应用平均数?

解 1. 平均数的计算与应用必须建立在同质基础上。马克思说: “平均量只能是种类相同的许多不同的个别量的平均。”不是同质的、不是同类型的观察值不允许合在一起求平均数。违背这一前提而算得的平均数是毫无意义的, 都是错误的。列宁指出: “将大小作坊混合在一起而得出‘平均’数字是完全荒谬的”。如在研究儿童生长发育时, 要在同年龄、同性别的基础上分别计算其平均身高、平均体重, 而不能把不同性别、年龄混在一起去算身高、体重两者的平均数。又如以某指标的均数作为考查某病是否治愈的标准时, 则要以病情等情况相似, 且确诊是该病的患者为对象。笼统的均数是完全虚构的数值, 它会造成假象, 掩盖本质, 以致会得出错误的结论。

2. 总均数与组均数的意义有所不同。科学的计算均数方法应以分组法为基础。总均数代表总的一般水平, 而组均数则常能更好地说明事物的局部特征, 以补充总均数。要善于用组均数与频数表来补充总的均数。

3. 分组要适当。用频数表法计算平均数时, 分组是否适当是影响计算结果的重要因素。若分组太粗, 组距过大, 组数甚少, 各组的组中值就不能有效的代表原来的观察值, 均数就会不准确。

在不等组距的分组时, 宜用 $\bar{X} = \frac{\sum f_x}{n}$ (X 为组中值) 式计算均数, 不能用