

# 实用学校卫生统计方法汇编

(第一辑)

《学校卫生》编辑部编印

1984年5月

# 实用学校卫生统计方法汇编

(第一辑)

《学校卫生》编辑部编印

1984年5月

# 编 辑 说 明

本汇编是帮助学校卫生工作者运用初步掌握的医学统计方法，解决学校卫生专业实际问题的参考资料。

本辑共收集44篇文章及有关附录，内容有：学校卫生统计资料的初步分析指标；生长发育及青春发育的资料收集、整理、分析与评价方法；儿少营养状况评价方法；疾病防治资料统计方法；学校保健组织与管理的统计研究方法以及进一步的统计研究方法等。最后附录有常用统计方法名词解释，常用医学统计方法流程图及医学生物统计方法公式用法指南。分两辑出版。

本书以普及对学校卫生专业实用的统计方法为目的，故未照搬一般卫生统计教科书内容。各篇文章虽独立，但编排由浅入深，尽量保持前后衔接与一定的系统性。每篇文章均简要介绍该方法的基本概念，并列出示范性实例演算步骤供读者参考。文字叙述力求通俗，以便读者通过自修方式学会运用。为便利广大读者，有些方法可能重复，但各作者阐述角度各异。未及汇编文章，有待今后陆续出版。

本书统计学符号主要参照上海第一医学院卫生统计学教研组《医学统计方法》和医科院卫生研究所等主编《卫生统计学》。本汇编曾邀请部分专家审阅，最后由阎承锐副教授初步校阅统一编排，原则上保持原著内容。但由于缺乏经验，水平有限，缺点错误难免，尚祈读者指正。

《学校卫生》编辑部《实用学校卫生统计方法汇编》编辑小组

1984年5月

# 目 录

## 编辑说明

学校卫生研究的统计方法问题……………阎承锐 李无为( 1 )

## 一、学校卫生统计资料的初步分析指标

- 1、相对数指标的应用问题……………李无为 阎承锐( 6 )
- 2、平均数指标的应用问题……………李无为( 12 )
- 3、正常值范围的估计……………李无为( 20 )
- 4、方差分析……………臧桐华( 32 )
- 5、形态机能发育资料的二元线性回归分析……………杨丽娟 李无为( 38 )
- 6、协方差分析在学校卫生中的应用……………臧桐华( 42 )
- 7、配对卡方检验的应用……………李为无 杨丽娟( 48 )
- 8、单一样本率的差别显著性检验……………胡克震 姜芝莉( 51 )

## 二、生长发育调查研究

- 1、学校卫生调查的抽样技术简介……………臧桐华( 82 )
- 2、生长发育测量资料的整理及统计方法……………荣光( 57 )
- 3、生长发育评价方法的选择( 附动态体型图 )……………姜芝莉 胡克震( 65 )
- 4、转盘式个体发育评价表简介……………西安市六中保健室( 70 )
- 5、用正态相关面法评价儿少身体发育之探讨……………陆守曾 马德胜( 74 )
- 6、儿童少年体格发育回归评价的计算图法……………马德胜( 79 )
- 7、排位次法在生长发育调查资料分析中的应用……………臧桐华( 108 )
- 8、逐步回归分析在生长发育研究中的应用……………臧桐华( 112 )
- 9、研究儿童群体生长发育的统计方法( 关于发育水平 )…汤旦林 许世瑾( 113 )
- 10、比较多群儿童身体发育指标的统计方法……………胡琳 马德胜( 131 )

## 三、青春发育研究方法

- 1、青春期发育水平评价及性发育的意义初探……………戴梅竞( 144 )
- 2、青春期形态、机能发育关系的复相关与多元分析……………周达生 戴梅竞( 149 )
- 3、青春发育各指标相互关系的偏相关及协方差分析……………周达生 戴梅竞( 155 )
- 4、青春期形态发育男女差别的理论研究……………戴梅竞等( 162 )
- 5、最优聚类分析—青春发育分期的数学方法……………周达生 戴梅竞( 169 )
- 6、月经初潮平均年龄统计方法……………马德胜( 177 )

7. 调查月经初潮平均年龄所需样本含量及误差的估算法初探……马德胜(192)

#### 四、儿少营养状况评价

1. “身高标准体重法”评价儿少营养状况的统计和使用方法……姜芝莉等(196)

2. 儿童现时营养状况的调查与评价……周宗棋(200)

3. 评价儿童营养状况的统计方法……詹绍康(205)

#### 附图 附表

(1)	式永平	附录四	……	1
(2)	姜芝莉	附录五	……	2
(3)	周宗棋	附录六	……	3
(4)	詹绍康	附录七	……	4
(5)	姜芝莉	附录八	……	5
(6)	姜芝莉	附录九	……	6
(7)	姜芝莉	附录十	……	7
(8)	姜芝莉	附录十一	……	8
(9)	姜芝莉	附录十二	……	9
(10)	姜芝莉	附录十三	……	10
(11)	姜芝莉	附录十四	……	11
(12)	姜芝莉	附录十五	……	12
(13)	姜芝莉	附录十六	……	13
(14)	姜芝莉	附录十七	……	14
(15)	姜芝莉	附录十八	……	15
(16)	姜芝莉	附录十九	……	16
(17)	姜芝莉	附录二十	……	17
(18)	姜芝莉	附录二十一	……	18
(19)	姜芝莉	附录二十二	……	19
(20)	姜芝莉	附录二十三	……	20
(21)	姜芝莉	附录二十四	……	21
(22)	姜芝莉	附录二十五	……	22
(23)	姜芝莉	附录二十六	……	23
(24)	姜芝莉	附录二十七	……	24
(25)	姜芝莉	附录二十八	……	25
(26)	姜芝莉	附录二十九	……	26
(27)	姜芝莉	附录三十	……	27
(28)	姜芝莉	附录三十一	……	28
(29)	姜芝莉	附录三十二	……	29
(30)	姜芝莉	附录三十三	……	30
(31)	姜芝莉	附录三十四	……	31
(32)	姜芝莉	附录三十五	……	32
(33)	姜芝莉	附录三十六	……	33
(34)	姜芝莉	附录三十七	……	34
(35)	姜芝莉	附录三十八	……	35
(36)	姜芝莉	附录三十九	……	36
(37)	姜芝莉	附录四十	……	37
(38)	姜芝莉	附录四十一	……	38
(39)	姜芝莉	附录四十二	……	39
(40)	姜芝莉	附录四十三	……	40
(41)	姜芝莉	附录四十四	……	41
(42)	姜芝莉	附录四十五	……	42
(43)	姜芝莉	附录四十六	……	43
(44)	姜芝莉	附录四十七	……	44
(45)	姜芝莉	附录四十八	……	45
(46)	姜芝莉	附录四十九	……	46
(47)	姜芝莉	附录五十	……	47
(48)	姜芝莉	附录五十一	……	48
(49)	姜芝莉	附录五十二	……	49
(50)	姜芝莉	附录五十三	……	50
(51)	姜芝莉	附录五十四	……	51
(52)	姜芝莉	附录五十五	……	52
(53)	姜芝莉	附录五十六	……	53
(54)	姜芝莉	附录五十七	……	54
(55)	姜芝莉	附录五十八	……	55
(56)	姜芝莉	附录五十九	……	56
(57)	姜芝莉	附录六十	……	57
(58)	姜芝莉	附录六十一	……	58
(59)	姜芝莉	附录六十二	……	59
(60)	姜芝莉	附录六十三	……	60
(61)	姜芝莉	附录六十四	……	61
(62)	姜芝莉	附录六十五	……	62
(63)	姜芝莉	附录六十六	……	63
(64)	姜芝莉	附录六十七	……	64
(65)	姜芝莉	附录六十八	……	65
(66)	姜芝莉	附录六十九	……	66
(67)	姜芝莉	附录七十	……	67
(68)	姜芝莉	附录七十一	……	68
(69)	姜芝莉	附录七十二	……	69
(70)	姜芝莉	附录七十三	……	70
(71)	姜芝莉	附录七十四	……	71
(72)	姜芝莉	附录七十五	……	72
(73)	姜芝莉	附录七十六	……	73
(74)	姜芝莉	附录七十七	……	74
(75)	姜芝莉	附录七十八	……	75
(76)	姜芝莉	附录七十九	……	76
(77)	姜芝莉	附录八十	……	77
(78)	姜芝莉	附录八十一	……	78
(79)	姜芝莉	附录八十二	……	79
(80)	姜芝莉	附录八十三	……	80
(81)	姜芝莉	附录八十四	……	81
(82)	姜芝莉	附录八十五	……	82
(83)	姜芝莉	附录八十六	……	83
(84)	姜芝莉	附录八十七	……	84
(85)	姜芝莉	附录八十八	……	85
(86)	姜芝莉	附录八十九	……	86
(87)	姜芝莉	附录九十	……	87
(88)	姜芝莉	附录九十一	……	88
(89)	姜芝莉	附录九十二	……	89
(90)	姜芝莉	附录九十三	……	90
(91)	姜芝莉	附录九十四	……	91
(92)	姜芝莉	附录九十五	……	92
(93)	姜芝莉	附录九十六	……	93
(94)	姜芝莉	附录九十七	……	94
(95)	姜芝莉	附录九十八	……	95
(96)	姜芝莉	附录九十九	……	96
(97)	姜芝莉	附录一百	……	97

# 学校卫生研究的统计方法问题

阎承锐

李无为

学校卫生研究的对象，一般都有一定的数量，因此必然涉及大量的统计问题。身高、体重、血压等个体差异（统计上称为“变异”）的出现则是由于许多内外因素随机交叉组合所致。此外，学校卫生调查研究都是根据有限的观测数据来推断总体情况，所得到的统计结论总是具有较大的准确性（可靠性），但又包含一定的错误率，因此，必须正确运用统计指标度量各种变异和误差才能够阐述有关事物或现象的规律性。

学校卫生工作的本身又是一个不断地调查研究过程。为了探索影响儿童青少年健康的各种因素，提出针对性的防治措施，必须经常地进行资料（特别要注重系统地累积资料）的收集、整理和分析工作。所以学校卫生工作与统计方法关系非常密切。根据学校卫生工作的实际，现就学校卫生研究统计方法的几个主要问题作几点说明。

## 一、课题与调查范围

学校卫生调查研究的课题，可大致分为两类。一类课题是以掌握基本情况为主，侧重描述；另一类课题是以探索为主，注重分析并提出假设进行验证。对于学校卫生工作基础薄弱，基本情况尚不清楚的单位，则应首先考虑弄清本地区学校卫生工作的基本情况。如进行学校卫生概况调查，可以了解学校卫生工作对象的特点、专业人员的配备、地区分布及演变等；又如通过对学生的生长发育调查、视力调查、健康状况调查、教学过程卫生调查、卫生习惯调查、营养调查……等，可以作出计量性统计描述。通过这些调查，不仅能了解本地区的现状，也为今后进行评价和比较积累了资料，并提供了一个可靠的对照基础。所以决不能忽视这些基础工作。

对于开展学校卫生工作已有一定基础的单位，则可以在现况（横剖面）调查的基础上开展深入调查研究。如进行中小学生病伤缺课因素调查，分析那些年级、何种疾病、什么时间、那些地区缺课多？究竟为什么多？进行身体形态发育测量（编制相关评价表），分析发育等级变动及其成因；在视力调查的基础上进行动态监测，分析那些年级、什么学习负担下视力低下多？与父母亲视力相关程度如何？在影响视力水平的多种因素中究竟那个是主要因素？还可以运用回归分析法研究身高、体重与血压、肺活量等，近视眼与影响近视的各因素的多指标间关系；还可以通过增减某个因素的方法进行实验性调查，如评价实行眼保健操在防治近视的作用；增加课间食对提高学习效果的作用；改善教室采光照明对保护视力的影响；比较少年运动员与非运动员形态或机能发育水平；比较大气污染区与未污染区中小学生有关疾病患病率（追踪观察）等。

学校卫生工作者，不仅要关心学生的身体健康，而且要重视学生的心理健康。所以，还可以进行青春期心理卫生，青少年吸烟，儿童智力水平等医学心理学和遗传学方

面的调查研究；也可以从学校卫生管理方面进行调查研究，如校医、保健老师的培训方法，医务室、保健室工作与装备要求，红十字少年活动，学校卫生服务工作质量评价方法等。总之，学校卫生工作实质上就是一个不断地进行调查研究的过程。

## 二、抽样调查与样本含量

统计学的调查方法分为全面调查与非全面调查两类。全面调查即普查，如欲发现某地区中小學生全部龋齿患者，则对该地区全体中小學生进行检查。非全面调查，虽可分为典型调查、重点调查等，但实质上都是抽样调查。即通过在总体中抽取一定数量的有代表性的样本进行调查来推论总体。我国1979年进行青少年儿童体质调查研究，并没有检查全国二亿多青少年儿童，而只是从中抽取一部分。全国共检查了十六个省市自治区7—25岁者二十多万人，就初步了解了某些规律和体质特点。总体相对大乃至接近数学上之无穷大时，除人口普查外一般不必要进行全面调查，因耗资费时，质量不易保证。学校卫生调查，一般为了探索事物规律，总结经验，收集资料准确、完整和及时，提倡有严密科学设计的抽样调查。

抽样调查有两个问题值得注意。一是抽样方法的选择问题；二是样本含量的大小问题。

抽样方法必须遵循随机化原则，保证样本对于总体的代表性。统计学中有抽签、随机数字表法的单纯随机抽样法；有如逢五抽一或隔三抽一的机械抽样法；有按集团抽取的整群抽样法和按照被研究事物或现象的有关特征进行分别按比例抽取的分层抽样法，以及分级多次应用随机抽样法等。其应用取决于对调查结果的准确度和精密度要求。例如，1980年某地中小學生形态发育调查，采用集团整群抽样方法，仅选取重点和非重点中小学校各一所，作为调查对象。而1981年他们编制《儿童少年形态发育相关评价表》时，则采用分层抽样方法，即在该地30所中小学校中，一中和一小以(一)班为调查对象，在二中、二小以(二)班为调查对象，余类推。(不够班数者则循环计数，如五小应以(五)班为调查对象，如只有三个班，则应抽(二)班)。用以推论该地区中小學生这个总体，其代表性明显地优于前法。

样本含量问题，虽然在习惯上有大样本与小样本之分，但首先要区别计量资料(对每个研究个体进行测量所得的数据，如身高、体重、等)还是计数资料(按每个研究个体的属性清点数目所得的数据，如性别、城乡等)。在用相对数作指标的计数资料中，如“夏天无”眼药水治疗近视的效果观察，观察30至50例以内，一般认为是小样本；50乃至100例以上，才是大样本。在用平均数作指标的计量资料中，如中小學生形态发育调查，按其性别、年龄别分组，7—17岁共须分作22组，按大样本要求，一般每组不得少于300例，则至少应调查6600人，才适应正态分布数学模型的估计要求。为了使样本含量估计适宜，卫生统计学专著上有估计样本含量的计算和查表法可供学习参考。(参见“正常值范围的估计”一文)

## 三、对照与条件控制

统计研究设置对照，就是在齐同的条件下，对进行研究的事物进行比较，从而客观

地判断某一种处理因素的优劣，或正确地推断两种属性的联系或因果。因此，没有对照，就不能作出科学的比较和判断。所以在进行设计时，不仅要有试验组，而且要有对照组。对照组与试验组之间除比较的因素外，其它一切条件尽可能一致。一般而言，自身对照的可比性比分组对照好；同一时期对照比不同时期好；本单位内对照比外单位好。例如，研究工业区大气污染对中小学生学习健康影响，试验组是大气污染地带的中小学生学习，对照组是大气未污染地带的中小学生学习，最好两组人群都居住在这个工业区内，经济水平、生活条件尽可能一致。有的课题也可以不设对照组，如某种疗法对高度近视的疗效观察。因为过去该病疗效极低的历史本身就是一个严格的公认的潜在对照。

常用的对照组设置有以下几种方法，都可根据不同情况应用于学校卫生调查。

**自身对照：**如“利福平治疗沙眼疗效观察”就是以治疗前和治疗后砂眼的变化作比较。

**相互对照：**如“某中学重点班与非重点班学生视力调查”，就是重点班与非重点班学生视力的相互比较。

**配对对照：**如“子女与父母视力相关调查”。就是把一名视力低下者随机选择性别、年龄、形态发育等级、学习环境、家庭经济状况相同的视力正常者配成一对来作为调查单位，探索其父母与子女视力遗传因素的关系。

**交叉对照：**如“雾视法与针刺疗法治疗近视的疗效观察”。先使用前法对甲组学生进行一个治疗程，用后法对乙组学生进行一个治疗程，然后两组疗法交换，交叉使用，进行疗效比较。但要特别注意先使用的处理因素，必须肯定没有后效应影响，否则难以判断效果。

**历史对照**以过去的情况作为对照或以别的单位和地区资料作为对照。由于对象、对比条件、环境等不尽相同，可比性较差，不是一个好办法。

对照组的建立，在对某个单因素研究时是适用的。但对于多因素研究，如形成近视的原因，运用单因素研究方法往往得到难以公认的结果，必须根据医学专业知识提出假设，把选择的多种因素科学组合，进行研究。本书介绍的多元回归、逐步回归、正交试验等方法，适合多因素分析，它可以对各种试验因素及其不同水平组合作出计量性估计。

#### 四、观察指标与数据处理

学校卫生调查与实验研究的结果能运用一个或几个观察指标来表示。因为观察指标是代表受试对象经过处理因素作用后出现的效应，是鉴定效果的尺度，所以观察指标的选择对调查研究工作质量关系很大。工作中应尽可能选用客观而敏锐的特异性观察指标。如在高血压防治研究中，用血压（特别是舒张压）作为观察指标，就比较客观，特异性也比较好；若用头昏、头痛、失眠等主观感觉作指标，不仅特异性差，而且带有难以克服的主观性。但是即使是客观指标，还有一个判断的主观性问题。如结核菌素试验强阳性儿童的胸片，骨龄调查的手骨X线片虽是一张客观的影像，但有时几个阅片者，就有不同的认识。所以在进行类似课题调查时，除要制定通用而明确的阅片标准外，还须采用交叉观察法或双盲观察法来避免主观性。



作为观察指标，要尽量精确。这个“精确”包括两层意思：一层是指标的正确性，即观察结果的可靠性如何；另一层是指标的精密度，即观察结果的重现性如何。准而不精，还可以粗略接近正确；精而不准，便没有实际意义。所以选择指标应注意客观性、特异性、可靠性和精密性。

在青春发育调查中，数据大多是计量的，一般都要计算平均数。但是统计学中平均数有多种计算方法，要根据资料分布的类型来选择。正态分布资料，如形态发育的身长、体重等指标，可用算术均数，也可以用中位数或众数作指标；偏态分布资料，如视力、营养状况，则只能计算中位数。此外还有百分位法、几何均数法、正态概率法、逆累计法、生存率法和寿命表法等用于各种青春期发育指标的计算。

在计算数据的正常值范围如血压、脉搏、肺活量、身长、体重、胸围等指标时，首先要确定选择80%，还是90%、95%乃至99%的正常值范围，还需选择单侧或是双侧界限，再根据正态性检验或 $\chi^2$ 检验、中位数检验等，确定分布类型后再酌选正态分布法、对数正态分布法或百分位数法进行计算。

对于两组间性别、年龄别等差异判断，应进行显著性检验，最常用的方法是t检验，还有更简便的符号检验等方法。随着统计研究的深入，青春期发育资料的正常值范围估计，还运用多元回归法，这比只有一个自变量的简单回归( $\hat{Y} = a + bx$ )更能缩小因变量对其估计值(预期值)的离散趋势，也就使正常值范围估计更为准确。

对于性征发育与形态发育的关系研究，还可运用协方差分析，就是把直线回归法与方差分析法结合起来，其目的是要把与Y值呈直线关系的X值化成相等后，再来检验各Y均数间差异的显著性，从而提高了统计效率。对于青春期发育多指标综合评价，本书还介绍试用聚类分析法。

在疾病调查中，观察指标大多是计数的，一般都需要计算相对数。如小学生龋齿患病率调查，中学生扁平足调查等都是把阳性体征人数作分子，受检人群作为分母计算其比例，再对性别、年龄别、地区别患病率差异作显著性检验，最常用的方法是U检验和 $\chi^2$ 检验。

在各种影响儿童少年健康因素的调查中，可应用相关研究法。研究比较简单的两个变量间的关系，即单相关(直线相关)；研究复杂的三个及以上变量之间的关系，即复相关和偏相关。如遗传对形态发育的影响，孪生子发育指标的比较等课题，都可以计算相关系数，并作显著性检验。如果采用配对调查设计，不仅可以用1:1的配对，而且可以作1:2的配对，都须计算 $\chi^2$ 值，以阐述其间相关的显著性。

总之，数据处理的统计方法颇多，应根据调查的目的及观测指标的特点酌情选用。

## 五、分析与判断结果

不论任何专业研究，总要包括受试对象，处理因素和处理效应这三个基本要素。因此，我们要根据专业知识提出假设，通过调查研究进行考核验证，力求阐述其因果关系。

调查资料经过统计处理，发现某种处理因素与某种效应联系有显著性叫统计联系。有的可能是因果联系，也有些并非因果联系，只是一种伴随现象，又叫相关关系。如“鸡鸣而破晓”说明鸡鸣与破晓两现象之间有联系，但并非因果联系。所以对于一般的

单因素研究，要认真设置对照，把回顾性调查结果作立论依据，解释要特别谨慎，最好是再进行前瞻性调查验证。动物试验结果更不能任意外推到人。

影响健康的致病因素很多，特别是对非传染病的病因研究建议从多因素、多病种、多学科的“三多”策略考虑，并从病因、环境和群体三方面去全面探索。例如，即使是象传染病那样十分明确的病因，也要进一步分析其必要条件与充分条件。如结核杆菌是结核病的致病微生物，是发生结核病的必要条件，但还不是充分条件。因为光有结核杆菌在人体内，并不能都发生结核病，还需要其它许多因素与结核杆菌协同作用，才能使人患病。所以中、小学生结核菌素试验的阳性者，并不都是结核病人。又如眼保健操作为中、小学生集体防治近视眼的措施实行已久，这就使人自然地联想到作为处理因素的眼保健操，作用于实验对象中，小学生是在什么处理标准上获得保护视力、预防近视的处理效应的。现在普遍存在的视力低下情况，究竟是不是眼保健操这个处理因素的质量标准没有掌握好？在这方面，1981年9月全国第一届儿少卫生学术会议上就有不同结论的报告。武汉医学院的报告指出：目前武汉市中、小学生做眼保健操水平，不足以产生防治学生近视眼的效果。山西省忻县地区卫生防疫站报告眼保健操对保护小学生视力有显著作用。山西医学院等单位报告，通过电子计算机进行九个因素的逐步回归分析的结果是：经常躺着看书和父母有无近视两因素对儿少近视发生有显著作用，没有把是否坚持眼保健操作为探索因素列入设计。从而说明设计与分析的角度不同，可以获得不同的结果。

所以当我们看到某因素与某现象呈高度相关的统计联系时，不能轻率地判定为因果关系。要作出因果关系的判断，须全面考虑以下条件：

1. 联系强度越大，成为因果关系的可能性越大。联系如能存在剂量反应曲线，如肺癌死亡率随吸烟支数增加，而呈线性上升，则说服力更强。
2. 联系较为恒定，能为其它人在不同时间、地区、环境及不同人身上所证实。
3. 联系具有特异性。这个特异性大大加强因果关系的成立。（如风疹与胎儿畸形。）
4. 联系能解释结果的分布特征，如疾病的空间与时间分布应与研究中因素分布一致。
5. 联系时间的先后，总是因在先，果在后，而决不能相反。
6. 因果联系应能言之成理，并符合已知的科学事实。

（详见苏德隆：联系与因果，中华预防医学杂志，1979年第13卷第2期106—113页）

（本文承上海第一医学院张国栋副教授审阅，谨谢。）

## 相对数指标的应用问题

李无为 阎承锐

医学上的数据资料,大体上可以分为两大类:一类是测量所得的记录,如身高、体重、血压等,称为计量资料(测量资料);另一类是清点数目所得的记录,如粪检蛔虫阳性人数、视力低下人数等,称为计数资料(属性资料)。计数资料本身是绝对数,为了描述和分析方便,常将有联系的绝对数相比计算出相对数。所以说相对数指标实质上是两个有联系的指标之比值。相对数指标在医学科研和学校卫生调查及卫生事业管理上应用极为广泛,数学计算很简单,但在实践中往往发生应用失误。所以在卫生统计方法教学中相对数指标往往有“易学难用”之说。

卫生统计学涉及卫生与统计两方面,不可单纯强调统计计算,必须具体地从数据的卫生学意义出发。所以在相对数指标的应用中,要特别注意指标的联系与对比。现归纳为十个问题,阐述如下。

### 一、各种相对数指标各有所用

相对数指标按其计算方法和说明问题不同可分为四类:即频率指标(也称率)、构成指标(也称构成比或比)、对比指标(也称相对比)和关系指标(也称关系比)。

1. 频率指标 是表示某现象实际发生例数与可能发生该现象的总例数之比,用来说明某现象发生的频率或强度。

$$\text{计算公式: 率} = \frac{\text{某现象实际发生例数}}{\text{可能发生该现象的总例数}} \times 100\%$$

如某中学初一年级新生400人体检发现视力低下者8人,那么,视力低下发生率 =  $\frac{8}{400} \times 100\% = 2\%$

2. 构成指标 是表示事物内部某一部分数值在全部数值总和中所占的比重,常以100作比例基数,所以也叫百分比。

计算公式:

$$\text{比} = \frac{\text{事物内部构成某一部分的数值}}{\text{事物各构成部分数值的总和}} \times 100\%$$

如某小学1980年因病伤缺课的720学时中,因传染病缺课240学时,传染病缺课占病伤总缺课学时的构成比 =  $\frac{240}{720} \times 100\% = 33.3\%$

3. 对比指标 是说明两个同类事物的数量之比,常以倍数或百分数表示。

$$\text{计算公式: 对比指标} = \frac{\text{某一事物的数量}}{\text{另一同类事物的数量}} \times 100\% (\text{或不乘比例基数})$$

如某市1979年高中毕业生体检合格率为84.84%，1980年为92.96%，其对比指标

$$= \frac{92.96}{84.84} = 1.096 \text{ (倍)},$$

其意义是1980年高中毕业生体检合格率是1979年的1.096倍，

或者说：1980年比1979年的体检合格率上升了9.6%。(或列式  $\frac{92.96 - 84.84}{84.84} \times 100\%$  也可。)

4. 关系指标 是说明两个有关但不同类事物的数量之比。

计算公式：关系指标 =  $\frac{\text{某一事物的数量}}{\text{另一有关非同类事物数量}} \times 100\%$  (或其它基数)。

如某矿山职工医院1980年医务人员共计126人，同年平均开放病床185张，其医务人员与开放病床之比 =  $\frac{185}{126} = 1.47$ ，即每名医务人员平均负担1.47张病床、也可以写作1:1.47。

## 二、要准确运用相对数指标

如发病率与患病率，虽只一字之差，意思迥然不同。发病率是指某种疾病新发病例(急性病每发一次，作为一个新病例，慢性病每年只统计一次)与同期年平均人口之比。

计算公式：发病率 =  $\frac{\text{某年某种疾病新发病例数}}{\text{同期年平均人口数}} \times 100\%$

它是描述某种疾病在一定时期一定人群中出现的频率，是一个动态指标；而患病率则是通过横断面调查了解特定时点上人群患病的断面情况，指的是现患，所以也叫现患率，它是调查时所发生病例及以前发病直至调查时尚未痊愈病例之和与断面调查人数之比。

计算公式：患病率 =  $\frac{\text{调查时所见病例数}}{\text{断面调查人数}} \times 100\%$ 。

它是一个静态指标。前者通过疾病登记搜集资料，多用于急性传染病的描述；后者通过医学检查发现病例，适用于慢性病的描述。

## 三、相对数指标要在相同条件下比较

在中小学校视力保护工作中，往往用视力低下率作为评价指标之一。但在两所学校或两个年级之间比较视力低下率高下来评价视力保护成绩时，首先要注意其本底水平的可比性即看两校或两个年级原来视力低下者中，高度近视者各占多少。因为高度近视在目前还没有较好的有效疗法，一般认为不是教学环境因素引起。所以尽管两校或两个年级原来视力低下者发生率相近，还要比较其视力低下人群中视力水平的构成。

再则，观察视力动态变化的视力监测，不宜采用断面调查方法，而应运用流行病学调查的队列(追踪)观察法，即把某学年入学的初一学生作为队列观察的成员，每学期定期进行一、二次视力监测。在分析比较时，每个学生的现在情况可以和自己的过去相比，每个班、每个年级也可以和本班或本年级的现在和过去比。如武钢三中1979~1981

学年入学学生初中三年的视力动态变化,就是运用队列观察法,男女学生的视力低下率、皆随年级上升而增加(表1)。

表1. 武钢三中1979年~1981年入学学生视力低下率(%)

	初 一	初 二	初 三
男	12.8	22.6	46.6
女	23.4	31.5	45.0

#### 四、观察例数过少或分母过小都不宜计算相对数

相对数指标本应能使我们掌握事物间客观存在的数量比例关系,但若观察例数过少,很难实现这种愿望。如某中药治疗肾炎并发尿毒症2例、其中治愈1例、则不宜将治愈率描述为50%。当然,这并不意味主张排斥医学科研中个别事物的分析研究,而是说这种条件下计算出来的相对数指标没有意义,倒不如保留绝对数的基本形式为好。因为抽样调查中观察例数愈多,相对数指标对总体的代表性愈好。又如在“某市居民区沙眼调查”一文中,分析不同年龄组发病率时,以51~60岁,61~70岁和70岁以上各年龄组检查38、31和18人,发现沙眼患者为24、19和10人,计算其患病率为67.2%、61.3%和55.6%来作出推论是不妥当的。因为用30~50例的小样本推论相当大的各年龄组总体,犹如坐井观天,片面性太大。

#### 五、时间因素对分子和分母有影响的情况

一种是分子和分母都随时间延长而积累,如病死率计算。随着时间的延长,积累的病例数增加,也就是分母数值增大而分子即病死的例数也可能随之有相应的增加,所以分子、分母比例水平不会发生太大的变化。

另一种情况是随着时间的延长,分母值相对稳定,而分子值随着时间延长而增加,如学生因病伤缺课率。这个指标分母是某一时期的学生在册人数,这个数值一般是稳定的,而分子是学生因病伤缺课学时,它与时间延长密切相关,观察时间越长,因病伤缺课学时积累就越多。因此,年度学生因病伤缺课率就会比季度、学期的因病伤缺课率高。

还有一种情况是随着时间的延长,分子和分母都不发生积累,在中小学生对流脑带菌情况不同季节的追踪观察时,带菌率计算的分母是每次受检人数、分子是受检者中出现阳性结果的人数,每次都是按人数作为计算单位。

#### 六、“合计”有不同的方法

不同含义的相对数指标,合计方法各有不同,应该注意区分,择宜选用。

构成指标的合计可以用相对数直接相加,但其和必须为100(表2)。凡是几个率并列,分子和分母都不重复计算时,分子和分母先分别相加,再重新计算合计的率(表3),绝不可将各个率相加而除之。

表2 1979年某公社各种疟疾的发病构成

	发病人数	百分比
恶性疟	68	70.10
间日疟	12	12.37
三日疟	17	17.53
合计	97	100.00

表3 1977年—1979年某县出血热病死率

	发病人数 (分母)	病死人数 (分子)	病死率 (%)
1977年	167	20	11.98
1978年	248	6	2.42
1979年	153	10	6.54
1980年	114	12	10.53
合计	682	48	7.04

凡是几个率并列、分母重复而分子不重复时，先将各个分母相加并求出商（平均分母），作为合计项的分母值，再将各个分子相加作为合计项的分子值，（表4）。凡是不同质的资料，不宜合计。如病人各种症状，合计症状出现率并没有实际意义，会导致错误的结论。

表4 1979年7—9月某矿职工菌痢发病率

	职工人数 (分母)	发病人数 (分子)	发病率 (%)
七月	6500	355	5.46
八月	6700	690	10.30
九月	6900	512	7.42
合计	6700	1557	23.24

### 七、根据率的含义选择分母

任何率的计算，分子都是属于分母当中的一部分，所以分母不同则意义不同。从某医院对62例可疑肝癌作超声波检查，并将结果与病理诊断结果对照（表5），说明用不同分母与分子组合计算出来的率，就有不同的含义（评价超声波诊断效果的相对数指标）。分别用检出率、漏诊率和误诊率来表示。

表5 某医院可疑肝癌患者超声波与病理诊断结果

		超 声 波 诊 断		
		肝 癌	非肝癌	合 计
病理 诊断	肝 癌	55	4	59
	非 肝 癌	3	—	3
	合 计	58	4	62

$$\text{检出率} = \frac{55}{59} \times 100\% = 93.2\%$$

$$\text{漏诊率} = \frac{4}{59} \times 100\% = 6.8\%$$

$$\text{误诊率} = \frac{3}{58} \times 100\% = 5.2\%$$

计算检出率和漏诊率是以经病理检查诊断为肝癌的59例作分母。因为只有当在确诊为肝癌的病例之中，才存在超声波检查的“检出”和“漏诊”的问题；而计算误诊率时，则应选择经超声波检查认为是肝癌的58例作分母，因为只有当在超声波检查认为是肝癌病例当中，才存在是否“误诊”的问题。所以不能选择全体受检人数62例作为分母来计算。

### 八、在一定条件下构成比例也可反映率的变化趋势

在一个地区和单位的一定时期内，人口变化不显著时，则可以从构成比上反映率的变化趋势。如从1979年某市菌痢逐月发病数和百分比上（表6），也可以看出5—10月是发病季节高峰。

在评价大气急性污染事件的危害时，往往用该地区呼吸系疾病、心血管疾病的死亡构成比的增加来描述危害程度和影响。

表6 1979年某市菌痢发病数与百分比

	月												合 计
	一	二	三	四	五	六	七	八	九	十	十一	十二	
发病数	23	19	12	36	51	58	66	98	60	59	26	28	536
百分比	4.3	3.6	2.2	6.7	9.5	10.8	12.3	18.3	11.2	11.0	4.9	5.2	100.0

### 九、不能单凭相对数指标的数字表面值作判断

如某钢铁公司中小学生视力减退率男性24%、女性16%的差异，究竟是抽样误差的影响，还是反映事物本质有差异呢？这就要求我们对样本推论总体时由于抽样误差而产生的不确定性给予度量，不过这种度量不是通过相对数指标的数字表面值，而是着眼于两者差异是否有显著性的显著性检验才能度量。又如，在“某市小学生沙眼调查”一文中，根据男生受检553人，患沙眼115人，患病率为20.8%；女生受检283人，患沙眼75人，患病率为26.5%，作者仅根据数值表面差别就作出沙眼患病率女生高于男生的结论，显然是不正确的。该例经过两个率的显著性检验，说明调查对象中男女学生沙眼患病率之间的差异没有显著性，现有差异来自抽样误差。所以单凭相对数表面值作为判断的依据，显然是不可靠的，也是不科学的。

### 十、总率（合计率）比较的标准化法

比较两个总率（合计率）时，如两组资料内部某项能影响统计指标水平的重要特征（如各年龄组调查人数）构成上有差别，则比较这两个总率是不合理的。

例如某大学1978年与1979年学生疟疾发病率比较（表7）。从总率（合计率）来看，1979年发病率高于1978年，但从学生来源分组看，不论来自低疟区和来自高疟区的学生，1978年发病率都高于1979年。因此比较两组资料总率来评价1979年的防治疟疾效果是不合理的。

表7 某大学1978年与1979年学生疟疾发病率比较

学生来源	1978年			1979年		
	人数	患者数	发病率(%)	人数	患者数	发病率(%)
来自低疟区	4000	40	10.0	1000	5	5.0
来自高疟区	1000	50	50.0	4000	120	30.0
合计	5000	90	18.0	5000	125	25.5

上述矛盾（假象）的出现，要归因于1978年和1979年招收学生来源人数构成大有差别。1978年来自低疟区学生人数远比1979年的多，来自高疟区的正好相反。为了正确比较，就要假设1978年和1979年学生来源人数有相同的构成，才不致因分组人数合计造成总率人为地上升或下降。然后根据这个共同的标准化的构成，计算标准化率，两组资料的总率才有可比性。当然，标准化的率只限于采用共同标准（人数）构成的两组（或几组）资料的互相比。由于选定的标准不同，所得数值各异。

本例以1978年学生来源人数构成为标准（当然也可以用1979年或其他别的标准人数构成），进行比较（表8）

表8 消除高低疟区学生人数构成差别影响后两年疟疾发病率比较

学生来源	1978年			1979年		
	人数	原发病率%	预期发病人数	人数	原发病率%	预期发病人数
来自低疟区	4000	10.0	40.0	4000	5.0	20.0
来自高疟区	1000	50.0	50.0	1000	30.0	30.0
合计	5000	—	90.0	5000	—	50.0

自上表算得1978年预期发病人数为90.0人（人数×原发病率），标准化发病率为

$$\frac{90.0}{5000} \times 1000\% = 18.0\%, \quad 1979\text{年预期发病人数则为}50.0\text{人，标准化发病率为} \frac{50.0}{5000} \times$$



1000‰ = 10.0‰, 可见, 1979年标准化发病率仍低于1978年, 与分别比较各年按高低病区学生来源分组的疟疾发病率结论一致。

(本文承上海冶金职业病防治研究所沈树春医师校阅。)

## 平均数指标的应用问题

李无为

学校卫生专业的许多研究课题, 诸如儿童少年生长发育的形态、机能、性征等指标; 健康状况调查的视力, 龋齿和血压等指标; 影响因素研究的教室卫生、课桌椅与营养状况等指标; 还有各种正常值范围研究, 都是以计算平均数(均数、均值)为基础。因此, 对于平均数指标的应用问题有加以研究的必要。

### 一、平均数的实质是一个表示集中趋势的位置量数

计量资料(测量资料)的每个观察单位都具有一个数值, 以数量大小表示, 称之为变量值或观察值。在一组测量资料中具有代表性的数值, 一般指在该组资料变量值集中处(中心位置)的数量, 即平均数。所以说平均数实质上是一个表示变量值集中趋势的位置量数。然而计量资料由于总体或者由于分组与计算方法不同, 出现不同形状的分佈曲线, 平均数不一定正好在数据资料的中心位置。在统计方法的专著上, 平均数可以计算算术均数、几何均数、中位数、众数和调和均数等多种指标。但在学校卫生专业上应用最广泛的是算术均数和中位数, 其次为几何均数。

算术均数的计算有直接法和频数表法两种。由于学校卫生研究多为大样本, 所以应用频数表法为多。直接法只适用于小样本资料。算术均数虽然容易被人们所理解, 但也特别容易受资料中特大值和特小值的影响, 使之相应地偏大或偏小, 未能确切的反映一组数据的平均水平, 它只是一组数据绝对和的平均而已。

中位数是频数次数的平均, 实质上是百分位数中的第50百分位数, 它是一种位次上的平均数。将全部变量值按由小到大顺序排列后, 将变量值个数除以2, 使之大于或小于第50百分位次的变量值个数相等, 这个变量值个数的一半就是中位数所在处。如9名8岁男孩身高值是122、123、123、124、125、126、127、128、128(厘米), 居中间位次是第5位, 即中位数是125, 若增加为10名, 增加139这个数值后, 居中间位次则在第5位与第6位之间, 即125和126之间, 则可计算出中位数为 $125 + 126 / 2 = 125.5$ 厘米, 代表性尚好。如果以算术均数平均表示,  $\bar{X} = 126.5$ , 则偏高。这是由于中位数是一种位次上的平均, 所以它不受资料中特大和特小极端值的影响。特别在屈光度、儿童营养状况的评价等计算时, 资料往往不呈正态分布, 这时应用中位数作平均数有突出的优点。

几何均数实质上是比例上的平均数, 主要适用呈等比关系的资料, 如血清学数据的处理。