



LXB

XUEYANJIUTONGJIFANGFASHILI

流行病学研究统计方法实例

● 章扬熙 编著

31-32

沈阳出版社

流行病学研究统计方法实例

章扬熙 编著

何观清 审阅

沈阳出版社

1989年·沈阳

责任编辑： 李孜孜

封面设计： 学 者

责任校对： 春 海

流行病学研究统计方法实例

章扬熙 编著

沈 阳 出 版 社 出 版 发 行

(沈阳市和平区十三纬路二段19号)

辽 宁 省 卫 生 厅 印 刷 厂 印 刷

开本 787×1092 1/32 字数200千字

印张 8.5

印数1—5000

1989年5月第1版 1989年5月第1次印刷

ISBN 7—80556—130—8/R·4 定价：3.85元

目 录

第一章	导言	1
第二章	流行病学研究中常用的统计方法 (一)	
2.1	流行病学研究中常用的统计指标	4
2.2	医学人口学研究中常用的统计指标	8
2.3	居民传染病漏报调查及其统计指标	10
2.4	两阶段整群抽样调查的统计推断	12
第三章	流行病学研究中常用的统计方法 (二)	
3.1	率的标准化	17
3.2	计量资料的分布及统计指标	21
3.3	假设检验	23
3.4	预防接种效果评价中的假设检验	26
3.5	参数估计	38
3.6	预防接种效果评价中的参数估计	39
3.7	相关与回归	49
第四章	专题性研究的设计	
4.1	专题性研究的种类	56
4.2	研究计划	68
4.3	抽样方法	71
4.4	样本大小	73
4.5	各种抽样的均数及标准误的计算	77
4.6	整群抽样的样本大小	86

4.7	病例对照研究的样本大小.....	87
4.8	定群研究的样本大小.....	88
第五章	爆发疫情的调查与统计分析	
5.1	爆发疫情调查的目的.....	90
5.2	步骤与方法.....	90
5.3	爆发疫情调查中常用的统计方法.....	94
5.4	爆发疫情调查与统计分析实例.....	97
第六章	疾病的预测预报方法	
6.1	疾病预测的依据.....	112
6.2	疾病预测的种类.....	113
6.3	定性预测.....	114
6.4	定量预测.....	136
第七章	病因研究中的统计分析	
7.1	病因研究的种类.....	145
7.2	团体的病例对照研究 (Mantel Haenszel 检验)	146
7.3	配比的病例对照研究.....	155
7.4	定群研究.....	160
第八章	防治措施效果评价的统计方法	
8.1	现场实验.....	173
8.2	现场观察.....	176
8.3	现场随访观察.....	184
8.4	措施的经济效益研究.....	188
第九章	寿命表法及其在流行病学研究中的应用	
9.1	完全寿命表.....	190

9.2	简略寿命表·····	194
9.3	寿命表中统计指标的方差与统计推断·····	198
9.4	去某死因寿命表·····	203
9.5	医学随访资料寿命表·····	204
9.6	医学随访资料寿命表中统计指标的方差 与统计推断·····	208
第十章 遗传流行病学双生子法的遗传度分析		
10.1	计数资料遗传度·····	216
10.2	计数资料遗传度的统计推断·····	217
10.3	计量资料遗传度·····	219
10.4	计量资料遗传度的统计推断·····	220
第十一章 Logistic回归模型及其在流行病学中的应用		
11.1	单因素的Logistic 回归模型·····	224
11.2	多因素的Logistic 回归模型·····	231
11.3	Logistic回归适用条件和在流行病 学中应用范围·····	237
11.4	非条件Logistic 回归模型·····	237
11.5	条件Logistic 回归模型·····	246
第十二章 Cox模型及其在流行病学中的应用		
12.1	Cox模型·····	253
12.2	Cox模型的参数估计·····	255
12.3	Cox模型的参数检验·····	260
12.4	危险指数与生存率·····	265

第一章 导 言

流行病学是研究疾病在人群中的分布及影响分布的因素，进而探索病因，阐明流行规律，拟订防治对策及措施并检验防治效果的一门科学。流行病学的研究对象是人群中的疾病现象，人群中的疾病现象是复杂多变的，并且是一个随机事件。对一个人来说，在特定的环境中，在一段时间内，某病可能发生，也可能不发生，发病可能轻，也可能重，可能痊愈，也可能死亡，但是，对于在特定环境中的整体人群来说，在某一时段内，某病的发病率、病死率是相对确定的。统计学是以随机事件为主要研究对象的，所以，流行病学的研究离不开统计学，统计学是研究流行病学的重要工具。

在疾病普查中，对疾病在人群中的三间（时间、空间、对象）分布应当怎样科学地进行描述和统计概括，需要借助于统计学。有时由于人力、物力和种种条件的限制，不可能对整个人群进行研究，只能抽取一部分，那么，如何抽样，又如何通过部分研究来推断总体，这也要靠统计学的帮助。

人群中疾病的动态分布受多种因素影响，这些因素作用的方向与程度如何，能否据此对某病进行预测预报，需要进行统计学的相关与回归分析。

对比分析是流行病学常用的方法。在抽样研究中，为了区别差异属于同总体的抽样误差，还是异总体的本质差别，

也需要统计学的帮助。

在流行病病因探索研究中，危险因子怎样确定，也需应用统计方法。比如，吸烟的人可能得肺癌，也可能不得肺癌，不吸烟的人也有可能得肺癌。那么，吸烟是不是发生肺癌的危险因子呢？危险度又有多大呢？这都要统计学的帮助。许多疾病是多因的，有些因子间还存在着交互影响，怎样综合考察多因素的作用，这属于统计学中的多因素分析和数学模型问题。

用统计方法对流行病防治措施的效果进行评价与探索病因有相似之处，前者考察保护因子的作用，后者考察危险因子的作用，二者均系干预因子，只是方向不同而已。在效果评价中除考察正效果外，还有副反应与经济效益问题。副反应与经济效益的研究，都有统计指标体系与统计推断问题，当然也离不开统计学的知识。

流行病学研究，主要是现场研究，包括现场调查和现场实验。现场情况是复杂的，除了研究因素以外，还存在着混杂因子。怎样在设计中使混杂因子均衡，或在分析中把混杂因子的作用与研究因素的作用相分离，也需要运用统计学的实验设计方法与分析技术。

流行病的疾病分布是动态的，人群中的疾病情况为随机事件，人群中的疾病过程亦是随机过程。而有关随机过程、时间序列的研究又属于统计学的范畴。

综上所述，流行病学与统计学的关系是十分密切的。对于流行病学工作者来说，应当掌握统计学知识与专业知识，在唯物辩证法的思想指导下，正确运用统计工具，以

便揭示疾病的危险因子与保护因子的作用，从数量上概括其流行过程，开展预测预报疾病发展趋势，判断预防效果，进行理论流行病学研究等工作。

以后各章将通过实例说明在流行病学研究中怎样具体应用统计方法。为了使读者能举一反三、触类旁通，在举例之前，对该统计方法的适用范围及公式均予以扼要介绍。

第二章 流行病学研究中常用的统计方法 (一)

2.1 流行病学研究中常用的统计指标

1. 发病(专)率 (specific morbidity rate): 表示一定时期(年、季、月等)内, 某人群中发生某病新病例的频率。

$$\text{发病(专)率} = \frac{\text{某年(期)内所发生的新病例数}}{\text{同年(期)平均人口数}} \times K \quad (\text{公式2.1})$$

式中 K 可为 100%, 1000‰, 10000/万, 100000/十万等。根据需要选用, 以下各公式同此。

2. 死亡(专)率 (specific mortality rate): 表示一定时期(通常为一年)内, 某人群中因某病而死亡者的频率。

$$\text{死亡(专)率} = \frac{\text{某年(期)内因某病死亡人数}}{\text{同年(期)平均人口数}} \times K \quad (\text{公式2.2})$$

上两式中平均人口数指年中人口, 可用 (年初人口 + 年末人口) ÷ 2 来计算。

3. 病死率 (fatality rate): 表示某病患者中因该病而死亡者的频率。

$$\text{某病病死率} = \frac{\text{因某病死亡人数}}{\text{同期某病病人数}} \times 100\% \quad (\text{公式2.3})$$

在以上三个基本率中，发病率常反映预防水平，病死率常反映医疗水平，而死亡率则为二者的综合。一个地区某病的死亡率低，或因发病率低，或因病死率低，或因二者皆低。所以，发病率与病死率是最基本的。三者的关系可以用下式表示。

某病死亡率 = 某病发病率 × 某病病死率

式中病死率用小数表示，即不乘100%。

4. 现患率 (prevalence rate): 这个统计指标常用于横断面调查，它表示某个时点上人群中存在某种疾病的频率。

$$\text{现患率} = \frac{\text{受检者中患某种疾病人数}}{\text{受检人数}} \times K \quad (\text{公式2.4})$$

对于慢性病长期监测点的资料，可用下式计算若干年平均现患率

$$\text{若干年平均现患率} = \frac{\text{若干年年末现患数的和} \div \text{若干年数}}{\text{若干年平均人口数}} \times K \quad (\text{公式2.5})$$

现患率与发病率不同，现患率的分子指调查时一切未愈的患者，而不管他是何年发病的都包括在内；而发病率的分子指在一定时期（通常为一年）内新发病例数。如果二者的人口数相同，发病率较平稳，同一地区某病的现患率与发病率的关系可粗略地用下式表示

$$\text{现患率}(Pt) = \text{发病率}(It) \times \text{平均病程}(D) \quad (\text{公式2.6})$$

〔例2.1〕 调查某市急性白血病患率为0.74/10万，

慢性白血病患率为 5.93/10万，二者的年发病率分别为 3.16/10万及 2.85/10万，求急性、慢性白血病的平均病程各多少？

将有关数据代入公式2.6得

$$\text{急性白血病平均病程 } D = \frac{P_t}{I_t} = \frac{0.74}{3.16} = 0.234 \text{ 年} \\ = 2.8 \text{ 月}$$

$$\text{慢性白血病平均病程 } D = \frac{5.93}{2.85} = 2.081 \text{ 年} = 25 \text{ 个月}$$

5. 罹患率 (attack rate)：表示一定人群中在较短时期新发病例的频率。较短时期通常可以月、周、日为单位，也可以一个流行期为阶段。用罹患率时，应注意暴露人口的准确性和时期的长短。

$$\text{罹患率} = \frac{\text{观察期间新病例数}}{\text{同期暴露人口数}} \times K \quad (\text{公式2.7})$$

式中K值常用100%或1000‰。

6. 携带率 (carrier rate)：表示某个时点人群中某种病原体携带者的频率。又称为带菌率、带毒（病毒）率等。

$$\text{携带率} = \frac{\text{受检者中携带某种病原体人数}}{\text{受检人数}} \times 100\% \quad (\text{公式2.8})$$

7. 感染率 (infection rate)：表示某个时点人群中受某寄生体感染的频率。受感染与否以体内有无寄生体或抗原或抗体为标志。

$$\text{感染率} = \frac{\text{受检者中某寄生体感染人数}}{\text{受检人数}} \times 100\% \text{ (公式2.9)}$$

8. 家庭二代续发率 (family secondary attack rate) : 表示家庭接触者接触原发病例后, 在一个潜伏期内续发某病的频率。

$$\text{家庭二代续发率} = \frac{\text{二代续发病例数}}{\text{家庭接触者人数}} \times 100\% \quad \text{(公式2.10)}$$

家庭二代续发率是反映原发病例对家庭成员威胁程度的指标, 也是考察疫区是否需要处理和评价处理效果的指标。另外, 家庭二代续发率比发病率通常要高, 因为家庭成员是密切接触者, 受感染机会比一般人群多。

〔例2.2〕 某年某地某病的发病率为 7%, 而家庭二代续发率为 1%, 续发观察期限为 7 天, 试求家庭接触者比一般居民发病的相对危险度?

首先推算家庭接触者的年发生频率为

$$1\% \div (7/365) = 52.14\%,$$

相对危险度为

$$52.14\% : 7\% = 74.5$$

这一结果表明, 家庭接触者感染该病的机会为一般居民的 74.5 倍。

9. 原发病例与续发病例之比 (primary case and secondary case ratio) : 这个指标反映各传染病疫区的扩延、发展情况。这里的续发病例包括家庭、同院、同单位等受原发病例感染而发病者。

10. 临床感染与亚临床感染之比 (clinical infection

and subclinical infection ratio)；它反映各种传染病在人群中显性与隐性感染的概貌，据此可从病例数的多少来推断感染人数的情况，洞察流行的分布全貌。

2.2 医学人口学研究中常用的统计指标

1. 出生率 (birth rate)：表示某年某地平均每千人口中的新出生人数，即活产数。

$$\text{出生率} = \frac{\text{某年出生人数}}{\text{同年平均人口数}} \times 1000\% \quad (\text{公式2.11})$$

2. 总死亡率 (mortality rate)：表示某年某地全人口中每千人中的死亡人数，它反映总的死亡水平。

$$\text{总死亡率} = \frac{\text{某年死亡总数}}{\text{同年平均人口数}} \times 1000\% \quad (\text{公式2.12})$$

3. 人口自然增长率 (natural increase rate of population)：表示某年某地每千人口中的增加人数（即出生数减去死亡数），它反映人口自然增减的情况。

$$\text{人口自然增长率} = \text{出生率} - \text{死亡率} \quad (\text{公式2.13})$$

4. 婴儿死亡率 (infant mortality rate)：表示某年某地每千名出生婴儿（不满周岁）的死亡数。

$$\text{婴儿死亡率} = \frac{\text{某年婴儿死亡总数}}{\text{同年出生总数}} \times 1000\% \quad (\text{公式2.14})$$

婴儿死亡率比总死亡率敏感，反映一个国家或地区的卫生、生活水平。

5. 新生儿死亡率 (neonatal mortality rate)：表示某年某地每千出生者中新生儿（未满一个月）的死亡数。

$$\text{新生儿死亡率} = \frac{\text{某年新生儿死亡总数}}{\text{同年出生总数}} \times 1000\text{‰}$$

(公式2.15)

婴儿死亡率在月龄上的分布呈反J型，越小的婴儿死亡率愈高。据统计，在婴儿死亡构成中，生后1小时内死亡的占总死亡人数的10%，生后1天内死亡的占20~25%，生后1个月内死亡的占50~70%，所以降低婴儿死亡率的重点在于降低新生儿死亡率。

6. 人口性别、年龄构成比 (constituent ratio of sex and ages of population)：先按性别分组，再按年龄分组计算人口构成比。年龄分组通常以5岁为组距，进而可以年龄为纵轴、以构成百分比为横轴，男于左侧，女于右侧，绘制人口性别、年龄构成图，即所谓“人口金字塔”。它可概括为静止型、增殖型和缩减型。目前我国呈缩减型，为计划生育的结果。

7. 围产儿死亡率 (perinatal mortality)：围产期指胎儿体重达到1000克，相当于足28周胎龄，直到出生后足7天。在这阶段的胎儿或婴儿称为围产儿。在这阶段死亡者为围产期死亡。围产儿死亡率用公式2.16计算，为

$$\text{围产儿死亡率} = \frac{\text{围产期死亡数}}{\text{活产 + 体重1000克以上 (或孕期满28周以上) 死胎、死产数}} \times 1000\text{‰}$$

(公式2.16)

8. 死产率 (stillbirth rate) 它是指出生体重在1000克及以上 (妊娠期在28周以上) 的总数 (包括活产、死胎和死产) 中，有多少胎儿死产，以千分率表示，用公式

2.17计算, 为

$$\text{死产率} = \frac{\text{28周及以上胎儿死产数}}{\text{活产} + \text{体重1000克以上(或孕期满28周以上)死胎、死产数}} \times 1000\text{‰}$$

(公式2.17)

2.3 居民传染病漏报调查及其统计指标

目前, 我国居民传染病漏报调查采用两阶段整群抽样的方法。在城区, 把调查地区划分为东、南、西、北、中五片, 每片抽查1~2地段, 每个地段抽查1~4个居委会。在农村, 也同样分成五片, 每片抽查1~2个乡, 每个乡抽查1~4个村。抽样人数与当地传染病总发病率有关, 要求能查出100例以上的法定传染病患者。

〔例2.3〕 某地法定传染病的总发病率为2000/10万, 求至少需要查多少居民?

我们希望查出100个传染病患者, 所以需查

$$\frac{100000}{2000} \times 100 = 5000 \text{人}$$

在进行漏报调查时, 要了解漏报的原因, 区分就诊漏报与漏诊漏报, 以从中寻求减少各种漏报的方法。在漏报调查中, 常用的统计指标有以下几种。

1. 某病漏报率: 计算公式为

$$\text{某病漏报率} = \frac{\text{查出某病漏报例数}}{\text{查出某病例数}} \times 100\% \quad (\text{公式2.18})$$

2. 估计发病率: 估计发病率为报告发病率 $\times$ 校正系数, 校正系数为 $\frac{1}{1 - \text{漏报率}}$, 所以估计发病率的计算公式为

$$\text{估计发病率} = \text{报告发病率} \times \left(\frac{1}{1 - \text{漏报率}} \right) \quad \text{公式(2.19)}$$

在应用上式时，漏报率用小数表示，估计发病率的单位与报告发病率的单位相同。

〔例2.4〕某地麻疹的报告发病率为90/10万，该地居民麻疹漏报率为30%，试求麻疹的估计发病率？

应用上式求得

$$\text{麻疹估计发病率} = \frac{90}{10\text{万}} \times \left(\frac{1}{1 - 0.30} \right) = 128.57/10\text{万}$$

3. 总漏报率：计算公式为

$$\text{总漏报率} = \left(1 - \frac{\sum \text{各病的报告病例数}}{\sum \text{各病的估计发病数}} \right) \times 100\% \quad (\text{公式2.20})$$

上式中估计发病数 = 估计发病率 × 总体人口数。总体是指调查地区的总体。总估计发病率的计算公式同估计发病率公式，其中报告发病率为总报告发病率，漏报率为总漏报率。

4. 漏诊率与迟报率的计算：漏诊率是指发生的传染病例中未就诊病例所占的比重，计算公式为

$$\text{漏诊率} = \frac{\text{未就诊传染病例数}}{\text{实际查出传染病例数}} \times 100\% \quad (\text{公式2.21})$$

迟报率是指超过传染病报告所规定的时限的报告病例数占同期已报告传染病例总数的比重，其计算公式为

$$\text{迟报率} = \frac{\text{迟报传染病病例数}}{\text{同期报告的总病例数}} \times 100\% \quad (\text{公式2.22})$$