

第三章 定性资料的统计描述

申旭波

shenxubo@zmc.edu.cn

公共卫生学院 预防医学教研室

本章内容

常用相对数：

对于定性变量，往往汇总具有某种属性的个体数；进而用相对数来描述和比较这类变量。

常用相对数指标：

介绍卫生统计业务中，有关指标的定义、性质、可能来源。

第一节 三类相对数

- 频率分布表：描述分布特征
- 相对数指标：统计描述和推断

一、频率与频率分布

表3-1 某医院2001年住院死亡病人中5类疾病的分布

疾病种类	死亡人数	百分比 (%)
恶性肿瘤	50	33.33
呼吸系统疾病	30	20.00
消化系统疾病	20	13.33
循环系统疾病	40	26.67
传染病	10	6.67
合计	150	100.00

- 样本中死因的频率分布：多个死因有多个频率，频率之和为1；
- 近似地反映同类医院中死因的概率分布：随机抽取一名死者，其死于各原因的概率。

- 从日常的观点看:目前事实的反映,死亡者中几种疾病的比重,或称构成比。

$$\text{构成比} = \frac{\text{某组成部分的观察单位数}}{\text{同一事物各组成部分的观察单位总数}} \times 100\%$$

构成比实为频率分布

例3-2 儿科医生共检查了210名儿童，发现有窦道乳牙尖炎患者136例。

- 儿童窦道乳牙尖炎的频率为
 $136/210=64.8\%$
- 不患此病的频率为
 $1-64.8\%=34.2\%$ 。

二分类变量，用一个频率64.8%就能描述频率分布。

频率型指标

- 频率型指标：最常见，近似地反映某一事件出现的机会大小，如发病概率、死亡概率等。

$$\text{频率型指标} = \frac{\text{某事件发生的个体数}}{\text{可能发生某事件的个体数}} \times K$$

K 是比例基数，取100%、1000‰、1万/1万和10万/10万

特点：分子是分母的一部分；
无量纲，在0~1范围内取值

二、强度

“强度”是流行病学、统计学术语

---- 单位时段内某现象发生的频率。如

$$\begin{aligned}\text{人时发病率} &= \frac{\text{新发病例数}}{\text{观察人时的总和}} \\ &= \frac{\text{新发病例数}}{\text{折合的观察人数} \times \text{单位时间}} \\ &\approx \text{单位时间内新发病的概率}\end{aligned}$$

例3-3 在某医院的院内感染调查中，5031个病人共被观察了127859人日（相当于127859人观察1日），其中有596人在医院发生感染。

观察人时数

$$\begin{aligned} &= 10 \text{ 日} \times 120 \text{ 人} + 9 \text{ 日} \times 240 \text{ 人} + 8 \text{ 日} \times 380 \text{ 人} + \dots \\ &\quad + 3 \text{ 日} \times 890 \text{ 人} + 2 \text{ 日} \times 920 \text{ 人} + 1 \text{ 日} \times 1200 \text{ 人} \\ &= 127859 \text{ 人日} \end{aligned}$$

$$\text{院内感染率} = \frac{596 \text{ 人}}{127859 \text{ 人} \times 1 \text{ 日}} = 0.0047 \text{ 人 / 人日}$$

平均每天有0.47%的病人将在医院发生感染。

一般地，

$$\text{强度型指标} = \frac{\text{某事件发生的个体数}}{\sum (\text{可能发生某事件的个体数} \times \text{时间})} \times K$$

其中比例基数 K 同上。

- 分母含有“时间”，并不是一般的频率，而是单位时间内的频率。

统计学和流行病学本质：频率强度或概率强度的近似值。

三、相对比

任何两个相关联的变量A与B之比

$$\text{相对比型指标} = \frac{A}{B}$$

- 相对于B的一个（或十个、百个、千个等）单位，A有多少个单位
- A和B可以是绝对数，也可是相对数
- A和B的量纲可以不同，也可相同，但A和B互不包含

四、相对数应用中需注意的问题

- (1) 防止概念混淆，即分析时不能以构成比代替率
- (2) 计算相对数的分母不宜过小,如：治疗4人，治愈率75%
- (3) 正确地合并估计频率（或强度）型指标
- (4) 资料的对比应注意可比性
 - a.观察对象要同质，即所比较组之间除研究因素外，其它重要的可影响研究结果的非研究因素应尽量相同.如检测手段、抽样方法、观察时间、地区、民族等尽可能一致。
 - b.其它影响因素在各组的内部构成是否相同，若不同应分组比较，或计算标准化率后再比较。
- (5) 样本率和构成比的比较，要作假设检验。

第二节 医学人口统计常用指标

一、医学人口统计资料的来源

日常工作记录；

统计报表；

人口调查（普查，抽样调查）。

二、描述人口学特征的常用指标

（一）人口总数（population size）

一个国家或地区在某一特定时间的人口数。

惯例：7月1日零时为标准时刻。

国际上两种统计人口数的方法：

- **实际制**，只统计标准时刻某地**实际存在**的人口数（包括临时在该地的人）；
- **法定制**，只统计某地的**常住**人口数。

从医学的角度来看，按实际制较好，如传染病的防治及计划生育管理等。

- 实际工作中，有时近似地
平均人口数 $\approx$ 人口总数

理论上，

平均人口 = 各时点人口数相加再除以总时点数

实际中，

平均人口 $\approx$ 相邻两年年末人口数相加再除以2

(二) 人口构成及其统计指标

指标	分子	分母	基数	类型
老年(人口)系数	≥ 65 岁人口数	人口总数	100%	频率型
少儿(人口)系数	≤ 14 岁人口数	人口总数	100%	频率型
总负担系数	≤ 14 岁人口数 + ≥ 65 岁人口数	15~64 岁人口数	100%	相对比
少儿负担系数	≤ 14 岁人口数	15~64 岁人口数	100%	相对比
老年负担系数	≥ 65 岁人口数	15~64 岁人口数	100%	相对比
老少比	≥ 65 岁人口数	≤ 14 岁人口数	100%	相对比
性别比	男性人口数	女性人口数	100%	相对比

三、生育和人口死亡的常用指标

- 世界卫生组织 “死亡” 定义：

“在出生后的任何时候，全部生命现象永远消失称为死亡”。

“胎儿死亡”，不应包括在生命统计的“死亡”内。

- 死亡统计资料

社会卫生状况和居民健康水平的基础资料

制订卫生工作计划、评价卫生服务效果的依据

（一）有关生育的常用指标

表 3-5 测量生育水平的统计指标

指标	分子	分母	基数	指标类型
粗出生率	同年活产数	同年平均人口数×1 年	1000/千	强度型(近似)
总生育率	同年活产数	同年 15~49 岁妇女数	1000/千	相对比型
年龄别生育率	同年某年龄组活产数	同年某年龄组平均妇女数×1 年	1000/千	强度型(近似)
总和生育率	是 15-49 岁年龄别生育率的总和			

(二) 死亡统计指标

表 3-6 人口死亡统计指标

指标	分子	分母	基数	类型
死亡率	同年内死亡人数	年平均人口数 $\times 1$ 年	1000/千	强度型(近似)
年龄别死亡率	同年某年龄组死亡人数	同年某年龄组平均人口数 $\times 1$ 年	1000/千	强度型(近似)
婴儿死亡率	同年 <1 周岁死亡人数	同年活产儿总数	1000/千	频率型(近似)
新生儿死亡率	同年 <28 天死亡人数	同年活产儿总数	1000/千	频率型(近似)
围生儿死亡率	同年围生期死胎数+死产数 + <7 天死亡人数	同年围生期死胎数+死产数+活产数	1000/千	频率型(近似)
5 岁以下儿童死亡率	同年 5 岁以下儿童死亡数		1000/千	相对比
孕产妇死亡率	同年孕产妇死亡数	同年活产儿总数	10 万/10 万	相对比
病因别死亡率	同年某原因死亡人数	同年活产儿总数	万	强度型(近似)
疾病病死率	同年某病死亡人数	同年平均人口数 $\times 1$ 年	10 万/10 万	频率型(近似)
死因构成比	同年某死因死亡数	同年患该病总数	万	频率型(近似)
		同年内死亡总数	100%	
			100%	

● 粗死亡率分母的近似计算：

假定： 当年每位去世者平均活了半年，
每位出生者也平均活了半年，

则有

$$\begin{aligned} & \text{当年全人口存活的总人年数} \approx \\ & \quad \text{年初人口数} \times 1\text{年} \\ & \quad - \text{去世人数} \times 0.5 \\ & \quad + \text{出生人数} \times 0.5 \end{aligned}$$

或

$$\begin{aligned} & \text{当年全人口存活的总人年数} \approx \\ & \quad \text{年终人口数} \times 1\text{年} \\ & \quad + \text{去世人数} \times 0.5 \\ & \quad - \text{出生人数} \times 0.5 \end{aligned}$$

- 上面两个等式相加除以2, 便有
$$\begin{aligned} & \text{当年全人口存活的总人年数} \\ &= \frac{1}{2}(\text{年初人口数} + \text{年终人口数}) \times 1\text{年} \\ &= \text{当年平均人口数} \times 1\text{年} \end{aligned}$$

或 用年中人口数近似地代替平均人口数

$$\begin{aligned} & \text{当年全人口存活的总人年数} \\ &= \text{当年年中人口数} \times 1\text{年} \end{aligned}$$

粗死亡率的定义式近似地写成

$$\text{粗死亡率} = \frac{\text{同年内死亡人数}}{\text{当年平均人口数} \times 1\text{年}} \times K$$

$$\text{粗死亡率} = \frac{\text{同年内死亡人数}}{\text{当年年中人口数} \times 1\text{年}} \times K$$

以上说明同样适用于年龄组死亡率、粗出生率和粗发病率等。

比较不同时期或不同地区的粗死亡率，如人口的年龄及性别分布不一致，应按性别、年龄标准化后作比较。

2. 反映死因分布及死因顺位的指标

- 比例死亡比（proportionate mortality rate, PMR），
又称死因构成

死亡原因	农村		城市	
	死亡专率	构成比	死亡专率	构成比
	(1/10 万)	(%)	(1/10 万)	(%)
呼吸系病	137.42	22.46	72.64	13.36
脑血管病	112.60	18.95	111.01	20.41
恶性肿瘤	105.36	17.73	135.59	24.93
心脏病	77.72	13.08	95.77	17.61
损伤与中毒	63.69	10.72	31.92	5.87
.....		...	...	...
合计		100%		100%

第三节 疾病统计常用指标

疾病统计（morbidity statistics）：疾病在人群中的发生、发展和流行分布的特点与规律。

为病因学研究、防治疾病和评价防治效果提供科学依据。

一、疾病统计资料的来源

1. 疾病报告和报表资料

- 国家规定的报告和报表：
法定传染病报告
地方病和寄生虫病报告
工矿企业职业病报告
- 某些部门规定的一些重要慢性病报告（结核病、恶性肿瘤、高血压病、冠心病）
- 死亡登记资料等

2. 医疗卫生工作记录

- 门诊医疗记录、门诊病历、住院病历、出院卡片等各种医疗记录。

3. 疾病专题调查资料

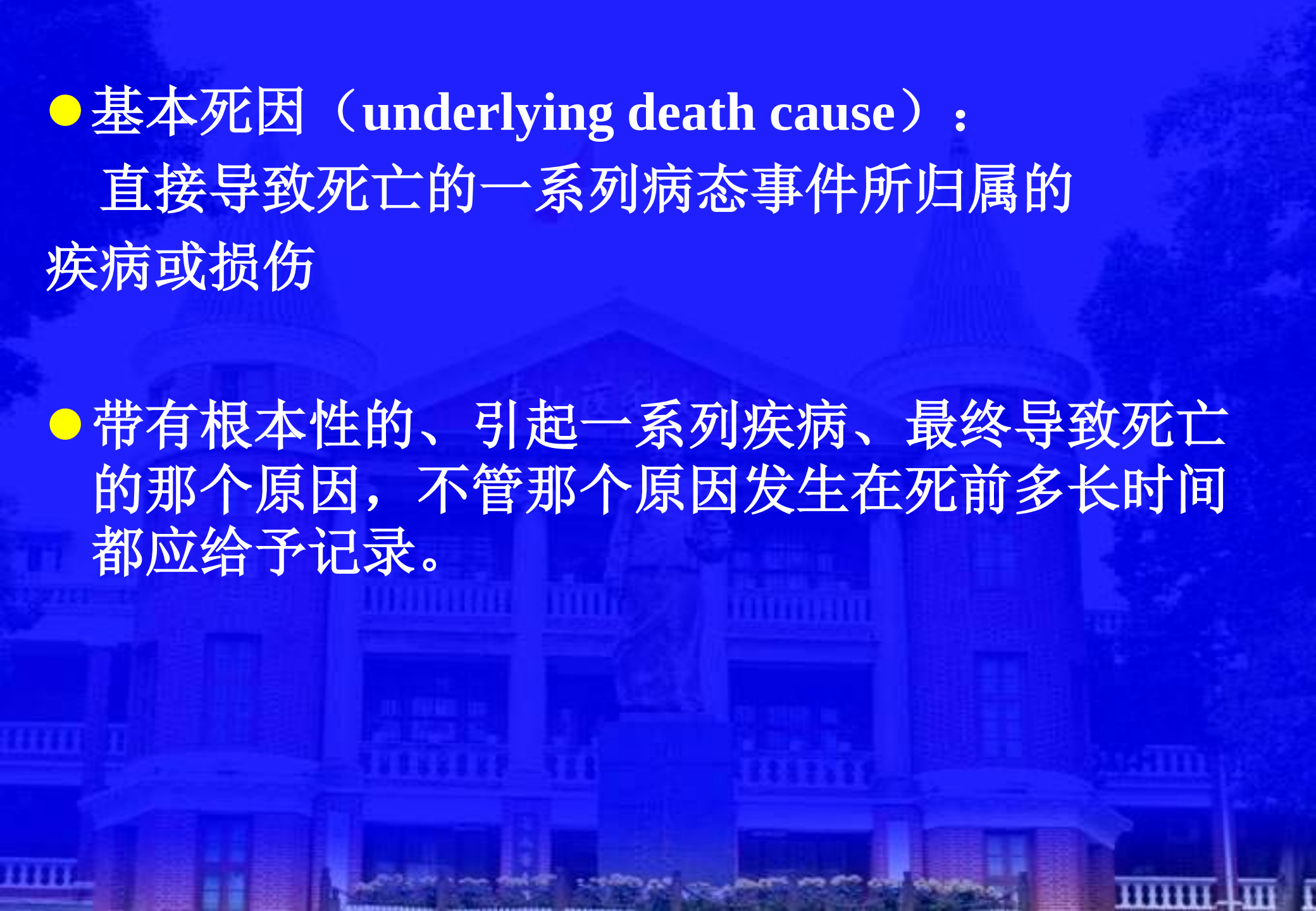
- 健康检查、疾病普查和疾病抽样调查等。
- 在资料的完整、正确可靠方面容易出现问题。

二、疾病和死因分类

- 疾病的命名（**nomenclature of diseases**）：标准化的术语
- 疾病分类（**classification of diseases**）：将具有共性的疾病归纳、分类。

区别：前者为疾病名称标准化，后者是统计研究需要。

- 1853年，国际统计学会着手编制了统一的疾病名称和死因分类，先是在欧洲使用。
- 1893年开始建立国际疾病和死因分类
(**International Classification of Diseases**, 简记为 ICD)
- ICD-10 : 1989年批准，第43届世界卫生大会上正式通过，1993年生效。

- 
- 基本死因（underlying death cause）：
直接导致死亡的一系列病态事件所归属的疾病或损伤
 - 带有根本性的、引起一系列疾病、最终导致死亡的那个原因，不管那个原因发生在死前多长时间都应给予记录。

三、疾病统计指标

疾病统计的单位：

- 病人：在观察期内，一个人是否转变为病人
- 病例：在观察期内，一个人每发生一次疾病就算一个病例。

指标	分子	分母	基数	类型
某病发病率	期内新发生的某病病例数	年平均人口数×1 年	10 万/10 万	强度 (近似)
时点患病率	时点现患疾病人数	检查人口数	10 万/10 万	频率
期间患病率	期内现患疾病人数	检查人口数	10 万/10 万	频率
残疾患病率	残疾患者人数	检查人数	100%	频率
治愈率	治愈人数	接受治疗人数	100%	频率
生存率	活满特定时期的人数	期初存活的人数	100%	频率

发病率（incidence rate, IR）：

“时期”通常用年或月。发病率的分母泛指一般的观察人年数,可由平均人口数×1年来近似。

新发病例数：新发生某种疾病，以第一次就诊为准。

用于探讨疾病的危险因素，评价疾病防治效果。

患病率（prevalence rate, PR）又称现患率：
受检人数中现患某种疾病的频率，通常用于慢性病。

- 时点患病率（point prevalence rate）
- 期间患病率（period prevalence rate）
主要用于对慢性病危害或防治效果的评价

在一定的人群和时间内，

（某病）患病率 = 发病率 × 病程

（某病）粗死亡率 = 发病率 × 粗病死率

- 类似于现患率的指标还有：砂眼检出率，寄生虫感染率，带菌率等。

生存率 (survival rate) :

病人能活到某一时点的概率

$$\text{生存率} = \frac{\text{活满特定时期的人数}}{\text{期初存活的人数}}$$

数据需通过随访获得，有直接法和寿命表法（详见本书后面有关章节）

- 治愈率 (cure rate) : 受治病人中治愈的频率
- 有效率: 受治病人中治疗有效的频率

第四节 粗率的标准化法

一、标准化法的意义和基本思想

2006 年上半年某医院普通外科甲、乙两组的手术效果比较

病情	甲			乙		
	治疗人数	有效人数	有效率 (%)	治疗人数	有效人数	有效率 (%)
轻	400	360	90	100	95	95
中	200	140	70	200	150	75
重	100	50	50	400	220	55
合计	700	550	79	700	465	66

哪个组的效果好?

甲、乙两县各年龄组人口及肺癌死亡率（1/10 万）

年龄组 (岁) (1)	甲县				乙县			
	人口 数 (2)	人口 构成 (3)	肺癌 死亡数 (4)	肺癌 死亡率 (5)	人口 数 (6)	人口 构成 (7)	肺癌 死亡数 (8)	肺癌 死亡率 (9)
0~	214833	0.5031	7	3.26	232677	0.6227	10	4.29
30~	65002	0.1522	6	9.23	36449	0.0976	7	19.20
40~	54320	0.1272	24	44.18	41478	0.1110	19	45.81
50~	43772	0.1025	50	114.23	34350	0.0919	42	122.27
60~	28340	0.0664	93	328.16	18657	0.0499	62	332.32
70~	20768	0.0486	62	298.54	10026	0.0268	32	319.17
合计	427035	1.0000	242	56.67	373637	1.0000	172	46.03

- 各年龄组肺癌死亡率：乙县>甲县；总死亡率：甲县（56.67/10万）>乙县（46.03/10万）。
- 为什么？

直接标准化法

表 16-2 直接法计算甲乙两县肺癌标准化死亡率率 (1/10 万)

年龄组 (岁)	标准 人口数	甲县		乙县	
		原肺癌 死亡率	期望肺癌死 亡人数	原肺癌 死亡率	期望肺癌 死亡人数
	N_i	p_i	$E_i=N_iP_i$	p_i	$E_i=N_iP_i$
(1)	(2)	(3)	(4) = (2) (3)	(5)	(6) = (2) (5)
0~	447510	3. 26	15	4. 29	19
30~	101451	9. 23	9	19. 20	19
40~	95798	44. 18	42	45. 81	44
50~	78122	114. 23	89	122. 27	96
60~	46997	328. 16	154	332. 32	156
70~	30794	298. 54	92	319. 17	98
合计	800672	56. 67	401	46. 03	432
	(N)		($\sum N_i p_i$)		($\sum N_i p_i$)

标准人口（第2列）：两县各年龄组的合并人口数 N

分别计算按照甲、乙两县的肺癌死亡率，标准人口会有多少人死于肺癌；合计值401，432分别是标准人口在甲、乙两县的期望肺癌死亡人数。

两县的肺癌标准化死亡率为：

甲县： $p_{\text{甲}}' = 401/800672 = 50.08$ (人/10万人年)

乙县： $p_{\text{乙}}' = 432/800672 = 53.96$ (人/10万人年)

标准化后：甲县<乙县，与分年龄组比较时的结论一致，解决了标准化前出现的矛盾。

一般公式：

已知标准人口年龄别人口数时：

$$p' = \frac{\sum N_i p_i}{N} \quad (16-1)$$

已知标准人口年龄别人口分布的频率时：

$$p' = \sum \left(\frac{N_i}{N} \right) p_i \quad (16-2)$$

直接法的思想：标准人口的总人口数为 N ，标准人口第 i 个年龄组的人口数为 N_i ，人口的频率为 N_i/N ，死亡率为 p_i 。**直接法标准化率实际上就是 p_i 的加权均数**，权重就是标准人口年龄的频率分布。甲乙两县采用统一的权重，其结果具有可比性。

间接法计算肺癌标准化死亡率

已知某年甲市因肺癌死亡280人，乙市为153人，两市数据如表16-5。试按间接法计算两市的肺癌标准化死亡率，并作比较。

间接标准化法

表 16-5 间接法计算甲乙两市肺癌标准化死亡率（1/10 万）

年龄组 (岁)	标准人口	甲市		乙市	
	肺癌死亡率	人口数	期望肺癌 死亡数	人口数	期望肺癌 死亡数
	p_i	n_i	$E_i=n_i p_i$	n_i	$E_i=n_i p_i$
(1)	(2)	(3)	(4) = (2) (3)	(5)	(6) = (2) (5)
0~	3. 80	368887	14	278225	11
30~	12. 81	64332	8	38944	5
40~	44. 89	55491	25	41480	19
50~	117. 76	42973	51	32330	38
60~	329. 81	26050	86	22671	75
70~	305. 25	11782	36	13485	41
合计	51. 71	569515	220	427135	188
			($\Sigma n_i p_i$)		($\Sigma n_i p_i$)

已知： 甲乙两市死亡总数及年龄组人口数

标准： 甲乙两县各年龄组合并的肺癌死亡率（第2列）

合计值220、188是甲乙两市的期望肺癌死亡人数：

甲市肺癌标准化死亡比 $SMR_{甲}=280/220=1.27$

甲市肺癌标化死亡率

$$p_{甲} = (51.71/10万) \times 1.27 = 65.67/10万$$

乙市肺癌标准化死亡比 $SMR_{乙}=153/188=0.81$

乙市肺癌标化死亡率

$$p_{乙} = (51.71/10万) \times 0.81 = 41.89/10万$$

一般公式：

$$SMR = \frac{r}{\sum n_i P_i} \quad (16-3)$$

$$P' = P \times SMR \quad (16-4)$$

间接法的思想： $\sum n$ 是按标准人口死亡率算得的被标准化组的期望死亡数， r 是被标准化组的实际死亡数； $r/\sum n_i p_i$ 称为**标准化死亡比**(standard mortality ratio, SMR)， $SMR > 1$ 表示被标准化组的死亡率高于标准人口死亡率； $SMR < 1$ 表示被标准化组的死亡率低于标准人口死亡率；**间接法标准化死亡率是对原总死亡率的校正，校正系数为SMR。**

三、应用标准化法的注意事项

1. 标准化法的应用范围很广：

- (1) 欲比较的两个人群内部的年龄分布不同；
- (2) 每个人群内部年龄别死亡率也各不相同。

发病率、病死率、患病率等。

2. 标准化死亡率并不是被标化组本身的某一种“真值”，而是用标准人口作为平台，对各被标化组进行的调整。标准化死亡率取值具有相对性。选择的标准人口不同，算得的标准死亡率也不同。

3. 标准组应选择有代表性的、较稳定的、数量较大的人群。

(1) 选择相互比较的人群之一作为标准组。一般选择人数较多者。

(2) 将各个比较组的数据合并作为标准组。

(3) 选择比较组之外的分布数据作为标准，如全国人口的年龄分布。

4. 标准化法常常用于对总体的标准化和比较。如是对样本资料进行标准化，其标准化死亡率的比较尚需作假设检验。

第五节 动态数列及其分析指标

表 3-13 某医院 1998-2005 年日门诊量动态变化

年份	指标	日门诊 人次	绝对增长量		发展速度%		增长速度%	
	符号		累计	逐年	定基比	环比	定基比	环比
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1998	a_0	2680	—	—	100.0	100.0	—	—
1999	a_1	3450	770	770	128.7	128.7	28.7	28.7
2000	a_2	3875	1195	425	144.6	112.3	44.6	12.3
2001	a_3	4488	1808	613	167.5	115.8	67.5	15.8
2002	a_4	5550	2870	1062	207.1	123.7	107.1	23.7
2003	a_5	5463	2783	-87	203.8	98.4	103.8	-1.6
2004	a_6	6600	3920	1137	246.3	120.8	146.3	20.8
2005	a_7	7602	4922	1002	283.7	115.2	183.7	15.2

一、绝对增长量

二、发展速度与增长速度

三、平均发展速度和平均增长速度

平均发展速度 = 环比的几何均数

$$= \sqrt[n]{\frac{a_1}{a_0} \cdot \frac{a_2}{a_1} \cdot \dots \cdot \frac{a_n}{a_{n-1}}} = \sqrt[n]{\frac{a_n}{a_0}}$$

平均增长速度 = 平均发展速度 - 1

小 结

1. 医学中的定性资料可通过频率分布进行全面的统计描述，也可以通过一些表现为相对数的统计指标从某个侧面进行描述。
2. 表现为相对数的统计指标大致有三种类型：
频率型、强度型和相对比型
3. 医学研究中关于定性变量的描述指标常涉及人口学特征、计划生育、疾病、死亡、残疾以及一些综合指标。这些统计指标在反映疾病负担、医疗质量和人群健康水平方面具有重要作用。

4. **死亡率的标准化**的基本思想就是寻找统一的标准组，计算标准化死亡率。若以**人口分布**为标准组则称直接标准化法，若以**死亡率分布**为标准组，则称间接标准化法。

5. 动态数列指标如发展速度和增长速度可描述事物在不同时期的发展水平和趋势。在实际工作中，要注意各种指标的正确计算方法和适宜的应用范围。