



成都中医药大学  
CHENGDU UNIVERSITY OF TCM

# 医学统计学

授课人：张杨、康娟娟



人民卫生出版社  
PEOPLE'S MEDICAL PUBLISHING HOUSE

**康娟娟 特聘副研究员 中医药创新研究院/交叉学科研究院**

**药用植物活性天然产物与生物合成研究团队**

- **研究方向：生物信息学**
- **办公室：温江校区弘景1号楼402**
- **电话：18200250974（微信同号）**
- **邮箱：[kangjuanjuan@cductcm.edu.cn](mailto:kangjuanjuan@cductcm.edu.cn)**
- **关键词：数据挖掘、机器学习、数据库**

## 第四章

# 定性数据的统计描述



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE



## 回顾

### 定量数据的统计描述

描述数据的两个重要的特征:

➤ 集中趋势:

算数均数

几何均数

中位数和百分位数

➤ 变异趋势:

✓ 按**间距**计算:

极差、四分位数间距

✓ 按**平均差距**计算:

方差、标准差、变异系数

- 在医学研究中,除了定量数据,还有如阴性和阳性、有效和无效、治愈和未治愈、生存和死亡,以及各种疾病分类等类型的定性数据。
- 对于这些数据的整理,往往先将研究对象按其性质或特征分类,再分别计数每一类的例数。



**Table 1 Estimated numbers of new cancer cases and incidence rates, overall, by sex and cancer type in China, 2016**

ICD10	Site	All			Men			Women		
		Cases	Crude incidence (1/10 <sup>5</sup> )	ASIRW (1/10 <sup>5</sup> ) <sup>a</sup>	Cases	Crude incidence (1/10 <sup>5</sup> )	ASIRW (1/10 <sup>5</sup> ) <sup>a</sup>	Cases	Crude incidence (1/10 <sup>5</sup> )	ASIRW (1/10 <sup>5</sup> ) <sup>a</sup>
C00-10, 12-14	Lip, oral cavity & pharynx	52200	3.78	2.43	36200	5.11	3.38	16100	2.38	1.48
C11	Nasopharynx	52000	3.76	2.51	37400	5.28	3.58	14700	2.17	1.42
C15	Esophagus	252500	18.26	11.13	184500	26.05	16.81	68000	10.07	5.60
C16	Stomach	396500	28.68	17.59	276300	39.02	25.14	120200	17.82	10.31
C18-21	Colorectum	408000	29.51	18.05	238500	33.68	21.65	169500	25.13	14.58
C22	Liver	388800	28.12	17.65	288800	40.78	26.65	100000	14.83	8.65
C23, 24	Gallbladder	55700	4.03	2.39	26400	3.73	2.36	29300	4.34	2.41
C25	Pancreas	100400	7.26	4.36	57000	8.05	5.14	43400	6.43	3.60
C32	Larynx	25700	1.86	1.17	23500	3.32	2.16	2200	0.33	0.19
C33, 34	Lung	828100	59.89	36.46	549800	77.64	49.78	278300	41.26	23.70
C91-95	Leukemia	85800	6.21	5.10	49400	6.97	5.77	36400	5.40	4.42
A_O	All other sites	173600	12.55	8.28	90900	12.83	8.81	82700	12.26	7.78
ALL	All sites	4064000	293.91	186.46	2234300	315.52	207.03	1829600	271.23	168.14

ICD-10, International Statistical Classification of Diseases 10<sup>th</sup> Revision; <sup>a</sup>ASIRW, age-standardized incidence rate by world standard

# 目录

- 第一节 常用相对数
- 第二节 医学中常用的相对数指标
- 第三节 相对数指标使用的注意问题



## 重点难点

掌握

率、构成比和相对比三种主要相对数类型、意义和计算，以及应用相对数时应注意的问题。

熟悉

医学中常用的几种相对数指标，如死亡率、发病率、患病率、病死率、相对危险度和比数比。

了解

标准化率的概念和计算方法。



## 第一节

# 常用相对数



人民卫生出版社  
PEOPLE'S MEDICAL PUBLISHING HOUSE





## 背景

- 在医学研究中，除了定量数据，还有如阴性和阳性、有效和无效、治愈和未治愈、生存和死亡，以及各种疾病分类等类型的定性数据。
- 对于这些数据的整理，往往先将研究对象按其性质或特征分类，再分别计数每一类的例数。
- 描述定性数据的数据特征，通常需要计算**相对数**。
- 包括：**率、构成比、相对比**等指标进行统计描述。



- **相对数定义：**两个有关联的绝对数之比，也可以是两个有关联统计指标之比。
- 相对数的性质取决于其分子和分母的意义，不同类型的相对数具有不同的性质。计算相对数的意义主要是把基数化作相等，便于相互比较。
- 举例说明：
- 某病用A法治疗100人，75人有效，B法治疗150人，100人有效。

若仅比较有效人数是不恰当的

通过计算其有效率分别为 $75/100=75\%$ ， $100/150=66.7\%$ ，进行比较疗效，才有意义。



## 一、率

- **率 (rate) : 表示在一定空间或时间范围内某现象的发生数与可能发生的总数之比, 说明某现象出现的强度或频度, 通常以百分率 (%)、千分率 (‰)、万分率 (1/万)、或十万分率 (1 / 10万) 等表示。**

$$\text{率} = \frac{\text{某事物或现象发生的实际数}}{\text{可能发生该事物或现象的总例数}} \times \text{比例基数}$$

- **公式中的“比例基数”通常依据习惯而定, 使算得的率小数点前面能保留至少1~2位整数。**





- 需要注意的是：率在更多情况下是一个具有时间概念的指标，用于说明某一段时间内某现象发生的强度或频率，如出生率、死亡率、发病率等，这些指标通常是指**在1年内发生的频率**。

- **例4-1** 某单位在2009年有3128名职工，该单位每年对职工进行体检，在这一年新发生高血压病人12例，则

$$\text{高血压发病率} = \frac{12}{3128} \times 1000\% = 3.84\%$$



## 二、构成比

- **构成比 (proportion) :** 表示某事物内部各组成部分在整体中所占的比重, 常以百分数表示。

$$\text{构成比} = \frac{\text{该事物内部某一组成部分的观察单位数 (例数)}}{\text{某事物内部的所有观察单位之和 (例数之和)}}$$

- **例4-2** 某医院某月住院病人数、死亡人数及统计指标如表所示。

某医院某月住院病人数及死亡人数统计

疾病类型	病人数	病死人数	死亡构成 (%)	病死率 (‰)
呼吸系统疾病	620	25	23.81	40.32
循环系统疾病	1030	35	33.33	33.98
消化系统疾病	540	20	19.05	37.04
恶性肿瘤	300	25	23.81	83.33
合计	2490	105	100.00	42.17



## 三、相对比

- **相对比 (relative ratio)**：是A、B两个有关联指标值之比，用以描述两者的对比水平，说明A是B的若干倍或百分之几，通常用倍数或分数表示。

$$\text{相对比} = \frac{A}{B}$$

- 这两个指标可以是性质相同，如不同时期的患病人数；
- 也可以是性质不同，如体重与身高的平方之比（体重指数，BMI）。

BMI=体重÷身高<sup>2</sup>。（体重单位：千克；身高单位：米。）

人体胖瘦程度	消瘦	正常值	超重
BMI数值	<18.5kg/m <sup>2</sup>	18.5-24kg/m <sup>2</sup>	>24kg/m <sup>2</sup> [5]



## 1. 两类别例数之比

- 例如我国2010年人口普查大陆31个省、自治区、直辖市和现役军人的人口中，男性人口数为686852572，女性人口数为652872280人，则

$$\text{男女性别比} = \frac{686852572}{652872280} = 1.052$$

即男性人口数是女性的1.052倍。





## 2. 相对危险度

- **相对危险度 (relative risk ,  $RR$ )** : 是流行病学前瞻性研究中常用的指标, 表示在两种不同条件下某疾病发生的概率之比, 反映暴露组发病或死亡的危险是非暴露组的多少倍, 说明疾病与暴露之间关联强度。若 $P_1$ 为暴露组的发病率;  $P_0$ 非暴露组的发病率, 则计算公式为

$$RR = \frac{P_1}{P_0}$$





## 例题

- **例4-3** 某地市区非吸烟女性饮酒者和不饮酒者的肺癌发病资料如表所示，试计算其相对危险度。

某地市区非吸烟女性饮酒者和不饮酒者的肺癌发病资料

暴露因素	饮酒与否	发病数	观察人年数	发病率 (1/10万人年)
	是	6	12965.2	46.3
	否	265	660291.4	40.1

$$RR = \frac{46.3}{40.1} = 1.15$$

说明该地市区非吸烟女性饮酒者的肺癌发病率是非吸烟女性不饮酒者的1.15倍。



### 3. 比数比

- **比数比 (odds ratio,  $OR$ )**：又称优势比，常用于流行病学中病例-对照研究资料，表示病例组和对照组中的暴露比例与非暴露比例的比值之比，是反映**疾病与暴露之间关联强度**的指标。

$$OR = \frac{P_1 / (1 - P_1)}{P_0 / (1 - P_0)}$$

$P_1$ 为病例组的暴露比例， $P_0$ 为对照组的暴露比例。



**例题**

➤ **例4-4** 母亲围孕期是否有发热或感冒病史与婴儿神经血管畸形关系的病例对照研究的资料如下表所示。试计算母亲围孕期是否有发热或感冒病史引起婴儿神经血管畸形的比数比。

母亲围孕期是否有发热或感冒病史与婴儿神经血管畸形关系

暴露因素

发热或感冒史	神经血管畸形组	对照组	合计
有	40 (a)	20 (b)	60
无	112 (c)	203 (d)	315
合计	152 (a+c)	223 (b+d)	375



**病例组**中的暴露比例与非暴露比例分别为

$$P_1 = \frac{a}{a+c} \quad 1 - P_1 = \frac{c}{a+c}$$

**对照组**的暴露比例与非暴露比例分别为

$$P_0 = \frac{b}{b+d} \quad 1 - P_0 = \frac{d}{b+d}$$

由OR计算公式可以得出

$$OR = \frac{P_1 / (1 - P_1)}{P_0 / (1 - P_0)} = \frac{[a / (a + c)] / [c / (a + c)]}{[b / (b + d)] / [d / (b + d)]} = \frac{ad}{bc}$$

本例  $OR = \frac{40 \times 203}{20 \times 112} = 3.63$

母亲围孕期是否有发热或感冒病史引起婴儿神经血管畸形的优势比为3.63。



## 四、标准化率

- **标准化率 (standardized rate)** : 比较两个不同人群的患病率、发病率、死亡率等资料时, 为消除其内部构成 (如年龄、性别、工龄、病程长短、病情轻重等) 对率的影响, 可以使用标准化率。
- **例4-5** 试计算甲乙两个医院的标准化率。

甲、乙两医院治愈率的比较

科室	甲医院			乙医院		
	入院人数	治愈人数	治愈率 (%)	入院人数	治愈人数	治愈率 (%)
内科	1500	975	65.0	500	315	63.0
外科	500	470	94.0	1500	1365	91.0
传染病科	500	475	95.0	500	460	92.0
合计	2500	1920	76.8	2500	2140	85.6



## 1. 标准化直接法的计算方法

$$P' = \frac{N_1P_1 + N_2P_2 + \cdots + N_kP_k}{N} = \frac{\sum N_iP_i}{N}$$

式中 $P'$  为标准化率,  $N_1, N_2, \dots, N_k$ 为某一影响因素 (如病型、年龄等) **标准构成的**每层例数,  $P_1, P_2, \dots, P_k$ 为原始数据中各层的率,  $N$ 为**标准构成的总例数**。上式也可写成:

$$P' = C_1P_1 + C_2P_2 + \cdots + C_kP_k = \sum C_iP_i$$

式中 $C_i=N_i/N$ 为标准构成比。



## 2. 标准构成的选取

➤ 标准化法计算的关键是选择统一的标准构成，选取标准的方法通常有下面三种：

- (1) 选取有代表性的、较稳定的、数量较大的人群构成为标准，如全国范围或全省范围的数据作为标准构成；
- (2) 选择相互比较的各组例数合计作为标准构成；
- (3) 从比较的各组中任选其一作为标准构成。



表 4-5 消除构成影响后两医院治愈率的比较

科室	标准构成人数	甲医院		乙医院	
		原治愈率(%)	预期治愈数	原治愈率(%)	预期治愈数
内科	2000	65.0	1300	63.0	1260
外科	2000	94.0	1880	91.0	1820
传染病科	1000	95.0	950	92.0	920
合计	5000		4130		4000

上例以各层例数的合计作为标准构成, 计算得到甲医院标准化后的总治愈率为

$$P'_{甲} = \frac{4130}{5000} \times 100\% = 82.6\%$$

乙医院标准化后的总治愈率为

$$P'_{乙} = \frac{4000}{5000} \times 100\% = 80.0\%$$

可见, 甲医院标准化后的总治愈率高于乙医院标准化后的总治愈率。



## 第二节

# 医学中常用的相对数指标



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE



## 一、死亡统计指标

1. **死亡率 (death rate)**：又称粗死亡率，表示某地某年**每千人口**中的死亡人数，反映当地居民总的死亡水平。

$$\text{死亡率} = \frac{\text{某年某地死亡人口数}}{\text{同年该地年平均人口数}} \times 1000\text{‰}$$

同年平均人口为年初人口和年末人口的平均值。

- 对不同死亡率进行比较时，应注意不同地区人口年龄或性别构成的影响。





2. 年龄别死亡率 (age-specific death rate) : 表示某地某年龄组每千人口中的死亡数, 它消除了人口年龄构成不同对死亡水平的影响。

$$\text{年龄别死亡率} = \frac{\text{某年某地某年龄组死亡人数}}{\text{同年该地同年龄别平均人口数}} \times 1000\%$$



3. **死因别死亡率** (cause specific death rate) : 表示某年每10万人中因某种疾病死亡的人数, 它反映各类病伤死亡对居民生命的危害程度, 是死因分析的重要指标。

$$\text{某病死亡率} = \frac{\text{某年某地某病死亡人数}}{\text{同年该地平均人口数}} \times 100\,000/10\text{万}$$

4. **死因构成** (proportion of dying of a specific cause) : 也称相对死亡比, 指全部死亡人数中, 死于某死因者占总死亡数的百分比, 反映各种死因的相对重要性。

$$\text{某病死亡构成比} = \frac{\text{因某种死因死亡的人数}}{\text{总死亡人数}} \times 100\%$$



## 二、疾病统计指标

1. **发病率 (incidence rate)**：表示在一定期间内，一定人群中某病**新发生**的病例出现的频率，是反映疾病对人群健康影响和描述疾病分布状态的一项测量指标。

$$\text{某病发病率} = \frac{\text{某时期某病新病例数}}{\text{同期间内平均人口数}} \times \text{比例基数}$$

需要注意的是，分母中所规定的平均人口是指可能会发生该病的人群。





2. **患病率 (prevalence rate) : 也称现患率, 表示某一时点某人群人口中患某病的频率, 通常用来表示病程较长的慢性病的发生或流行情况。**

$$\text{某病患病率} = \frac{\text{某地某时点某病患者例数}}{\text{该地同期间内平均人口数}} \times \text{比例基数}$$

**以上比例基数可为100%、1000‰、10000 / 万、100 000 / 10万, 实际中患病率的分母通常为调查的总人数, 分子为患病的人数。**



3. **病死率 (case fatality rate)** : 表示某期间内, 某病患者中因某病死亡的频率, 表明该疾病的严重程度和医疗水平等, 多用于急性传染病。

$$\text{某病病死率} = \frac{\text{某期间因某病死亡人数}}{\text{同期该病的患病人数}} \times 100\%$$

4. **治愈率 (cure rate)** : 表示接受治疗的病人中治愈的频率。

$$\text{治愈率} = \frac{\text{治愈病人数}}{\text{接受治疗病人数}} \times 100\%$$

### 第三节

## 相对数指标使用的注意问题



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE





- 1. 不要把构成比与率相混淆：**构成比只能说明某事物内部各组成部分的比重和分布，不能说明该事物某一部分发生的强度与频率。
- 2. 使用相对数时分母不宜过小：**分母过小时相对数不稳定。在观察例数较小时，最好直接用绝对数表示，以免引起误解。
- 3. 正确计算合计率：**对观察单位数不等的几个率，其合计率的计算不能将各个率相加求其平均值，而应该用各率的分子之和除以分母之和来计算。
- 4. 注意资料的可比性：**比较的资料应是同质的，即除了要比较的处理因素外，其他条件应尽量相同或相近。
- 5. 样本率或构成比存在抽样误差：**比较各组的样本率或构成比的差别需作假设检验(第九章)。



## 本章小结

1. 定性资料的变量形式包括多分类变量和两分类变量，这些变量可以通过相对数对其特征进行统计描述。
2. 常用的相对数指标有率、构成比和相对比。相对数的性质取决于其分子和分母的意义，不同的相对数其指标的定义和结果解释不同，使用时注意各自的用法和意义。
3. 标准化法的目的是消除重要因素的构成不同对粗率比较的影响，实际中可以选择统一的“标准”计算标准化率。



## 本章小结

4. 医学人口统计和疾病统计中有关人口死亡和疾病的一系列指标是定性资料统计描述的应用。这些统计指标在确定卫生政策、了解人群健康水平和评价卫生工作效果、反映疾病负担和医疗质量方面等具有重要作用。 $RR$ 值和 $OR$ 值则可以衡量相对风险。





## 课后习题二

### 1. [参考答案]

不正确。因为此百分比是构成比，不是率，要知道男女谁更易患病，需得到1290名职工中的男女比例，然后分别计算男女患病率。

### 2. [参考答案]

不正确。此为构成比替代率来下结论，正确的计算是用各年龄段的死亡人数除以各年龄段的调查人数得到死亡率。

### 3. [参考答案]

不正确，此为构成比替代率来下结论，正确的计算是用各型肝炎的新病例数除以同时期内可能会发生该病的人群人口数得到发病率。





## 4、两煤矿工人尘肺总患病率

工龄	总人数	甲患病率	甲患病人数	乙患病率	乙患病人数
0~<6	15018	0.86%	129.1548	0.2%	30.036
6~<10	6190	3.92%	242.648	0.42%	25.998
10~30	3556	12.43%	442.0108	11.54%	410.3624
所有年龄	24764		813.8136		466.3964
		3.28%		1.88%	





5. 现有甲乙两地某月居民的粗死亡率资料如下表所示,试比较两地死亡率水平高低。

甲乙两地某月居民粗死亡率比较

年龄(周岁)	甲地区			乙地区		
	死亡数	人口数	死亡率(‰)	死亡数	人口数	死亡率(‰)
0 ~	20	1000	20	180	6000	30
25 ~	120	3000	40	150	3000	50
65 ~	360	6000	60	70	1000	70
合计	500	10 000	50	400	10 000	40





## 5、甲乙两地某月居民死亡率

年龄段	总人数	甲死亡率	甲人数	乙死亡率	乙死亡人数
0~<25	7000	20‰	140	30‰	210
25~<60	6000	40‰	240	50‰	300
60~	7000	60‰	420	70‰	490
所有年龄	20000		800		1000
		4%, 40‰		5%, 50‰	



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE

谢谢观看