

第一章 统计图表

——用图形表示数据（分布）的特点

数据的表达与描述的方法

- 描述统计（**descriptive Statistics**）主要研究如何整理实验或调查得来的大量数据，描述一组数据的全貌，表达一件事物的性质。
- 描述数据的常用方法：
 - 表格法
 - 图示法
 - 数值指标法

一、统计表——概念与目的

- 概念

统计表是用于表达被研究对象的特征、内部构成及研究项目分组之间的数量关系的表格。

- 目的

使人能一目了然地看出被说明事物的特点、关系、发展趋势等，同时便于分析运算。

一、统计表——一般统计表的结构

• 统计表的结构：

- 表号
- 名称（标题）
- 项目（纵标目、横标目）
- 数字
- 数据来源
- 线条
- 表注

The diagram illustrates the structure of a statistical table with the following components labeled:

- 表号** (Table Number): 表1
- 标题** (Title): 某县高考被录取学生家长职业比较表
- 顶线** (Top Line): The line above the header row.
- 表线** (Table Line): The line separating the header row from the body rows.
- 标目** (Label): Points to the vertical header '家长职业'.
- 标目** (Label): Points to the horizontal header '高考录取人数'.
- 标目** (Label): Points to the horizontal header '所占比例(%)'.
- 数字** (Number): Points to the value '128' in the body.
- 底线** (Bottom Line): The line below the last row of the table.
- 表注** (Table Note): *表内数据来自县招生办

表1 某县高考被录取学生家长职业比较表		
家长职业	高考录取人数	所占比例(%)
农民	128	43.7
教师干部	63	21.5
工人	22	7.5
其他	80	27.3
总计	293	100

*表内数据来自县招生办

一、一般的统计表——统计表的结构

表号 标题（包括何时、何地、何事）

横标目的 总 标 目	总 标 目		总 标 目（单位）			
	纵 标 目	纵 标 目	纵 标 目	纵 标 目	纵 标 目	纵 标 目
横 标 目	×××	×××	××.	××	××.	××
┆	××	××	×.	××	×.	××
┆	┆	┆	┆	┆	┆	┆
┆	┆	┆	┆	┆	┆	┆
合 计	×××	×××	××.	××	××.	××

备注：

一、一般的统计表——统计表的种类

表2 某校初一各班数学成绩比较表

班次	平均成绩
一	69
二	72
三	75
总平均	73.5

- 统计表的种类

- **简单表**(simple table): 只按一个特征或标志分组。如表2。

- **复合表**(combinative table)：按两个或两个以上特征或标志结合起来分组。如表3。

表3 某校教师学历统计表

最后学历	毕业时间	人数	百分比
大学	66年以前	21	16
	83年以后	11	8
大专	66年以前	74	54
	83年以后	29	22
合计		135	100

一、一般的统计表——编制要求

- 编制统计表的要求

- 标题的措词要简明扼要。表号与标题之间留下一个汉字的空格。
- 内容清楚，标目分类要准确，以能说明问题为主。
- 数据是统计表的语言，又称统计指标。要求用阿拉伯数字，准确、整齐，统一。单位一般归放在标目中。
- 线条不要太多，表的上下端有顶线与底线，左右两边不封，合计项须用线条与其他项目分开。

一、一般的统计表——常见错误

- 线条过多、数字的小数点位数不一致、单位在数字之后。
- 表中数据的含义未表达清楚，令人费解。

二、次数分布表

- 次数分布 (**frequency distribution**) 一组数据的分布情况。
- 如：每一个观测值出现的次数，或者是数据在每个数值区间内的散布情况。
- 次数分布表和次数分布图就是各种次数分布的列表形式和图示形式。

二、次数分布表——简单次数分布

- 例：某公司人力资源部使用调查问卷对某部门员工实施调查。其中有一道题是：“你认为本部门现任主管尽职的程度如何？”

表2-1 员工对部门主管尽职程度的评价

评定等级	人数
①非常不尽职	9
②不尽职	30
③不置可否	10
④尽职	25
⑤非常尽职	6
总计	80

二、次数分布表——分组次数分布表

- 对于连续数据或数据点很多时，先划分出若干分组区间，分别统计各个区间中包含的数据个数，并用列表的形式呈现出来。

编制分组次数分布表（grouped frequency table）的步骤：

（1）求**极差**（range），又称为全距。

（2）决定**组数**、**组段**和**组距**：根据研究目的和样本含量n确定。通常分10-15组，为方便，组距可参考极差的1/10，再略加调整。

$$K = 1.87(N - 1)^{\frac{2}{5}} \quad \text{组距} = \text{极差} / \text{组数}$$

二、次数分布表——分组次数分布表

(3) 列出组段：第一组段的下限略小于最小值，最后一个组段上限必须包含最大值，其它组段上限值可忽略。

(4) 划记计数：用划记法将所有数据归纳到各组段，得到各组段的频数。

(5) 编制成表

二、次数分布表——相对次数与累积次数分布表

- 将次数分布表中各组的实际次数转化为相对次数，即用次数比率（ f/N ）或次数百分比（ $f/N \times 100\%$ ）来表达数据分布情况，就可制成相对次数分布表。
- 累积次数是把各组的次数由上而下，或由下而上累加。
- 用累积次数表示的次数分布称为累积次数分布（**cumulative frequency distribution**）。
- 累积次数也可转化为相对累积次数。或者说是将相对次数进行累加，得到的次数称为相对累积次数。最后一组的相对累积次数应等于？

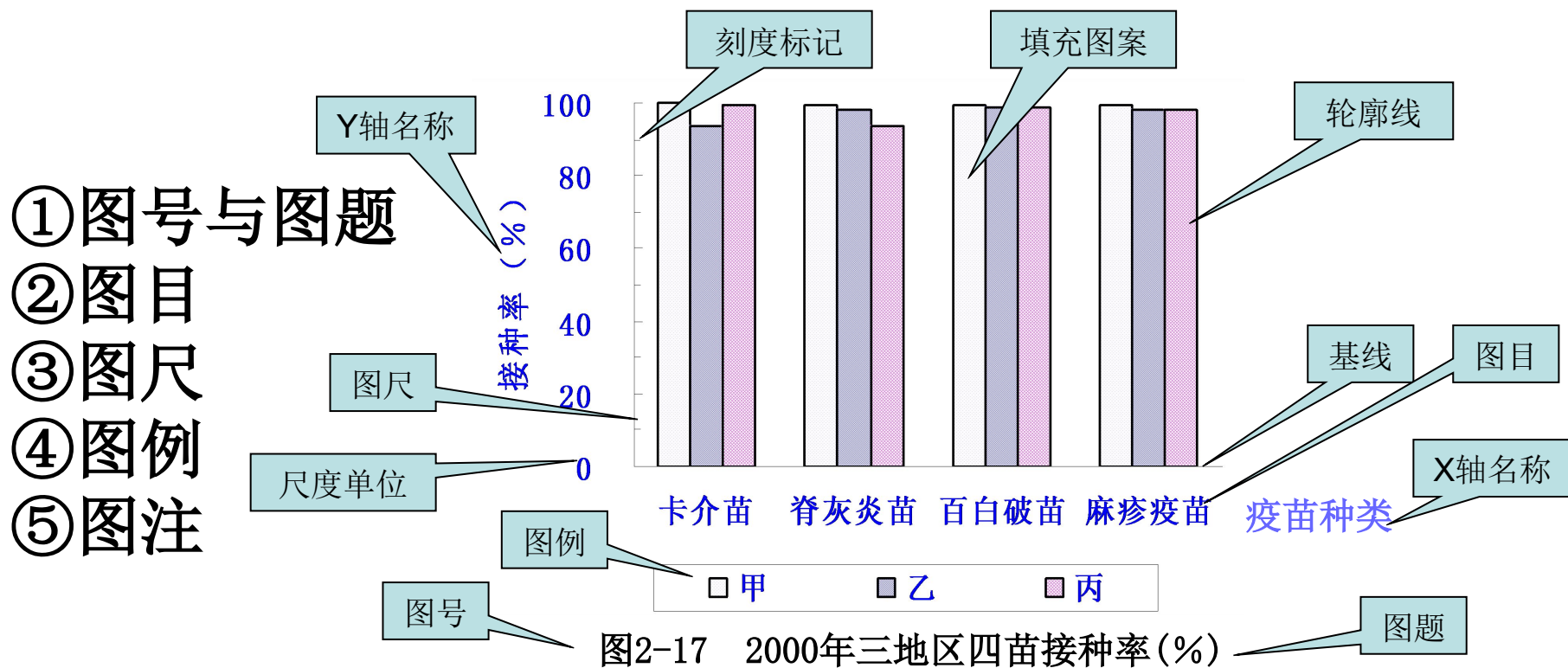
三、常用的统计图

- 统计图(statistical chart 或 statistical graph) 是用点、线、面、体等几何图形，直观形象地表达、描述数据或结果。
- 根本目标
 - 形象化的方式把事物的特性、规律显示出来。
 - 使人能获得全面与深刻的直观形象。
 - 便于说明问题与比较。

三、常用的统计图——统计图的结构

表2-17 2000年三地区四苗接种率(%)

地区	接种率(%)			
	卡介苗	脊灰炎苗	白百破苗	麻疹疫苗
甲	99.72	99.20	99.24	99.12
乙	93.50	98.20	98.70	98.20
丙	99.50	93.90	98.70	98.20



三、常用的统计图——统计图的种类

• 统计图的种类

- 条形图(Bar chart)
- 圆形图(Pie chart)
- 百分条图
(percentage chart)
- 线图 (line chart)
次数多边形图
累积次数分布图
- 直方图 (histogram)
- 箱图(boxplot)
- 茎叶图(stem-and-leaf display)
- 散点图(scatter diagram)
- 气泡图
- 统计地图(statistical map)
- 雷达图(radar graph)

三、常用的统计图—— 绘制统计图的要求

- 要求：
 - 根据数据和目的选择合适的图形
 - 图形所表示的面积或距离要比例适当
- 分布图的分析
 - 峰点 (peak) 研究
 - 对称性 (symmetry) 研究
 - 形状 (shape) 研究：
 - 年龄与死亡率 U 婚姻次数与人数 L
- 常见错误：
 - 坐标轴上的刻度值标得不符合数学原则。
 - 数据含义表达不清楚。

表 各种统计图适用的资料类型与分析目的

图类型	资料性质	分析目的
1. 条图	横轴为间断独立的分组	直条长短表达统计指标大小
2. 饼图	构成比	条段面积表达构成
3. 圆图	构成比	圆的扇形面积表达构成
4. 线图	横轴为连续变量	统指随变量变化的趋势
5. 半对数图	横轴同上，纵轴取对数	统指随变量变化的速度
6. 直方图	横轴为分组，纵轴为频数	直条面积表达频率
7. 直方图	直方图的变形	同上
8. 散点图	两个变量	用点的密集程度和趋势反映两变量关系
9. 统计图	地域性资料	点的疏密、颜色的深浅等说明疾病地域分布

统计表 统计图

标题	简明扼要给出图表的基本内容	
	表上端中央	图下方正中
标目	如有度量单位，应标注	
横	一般为叙述事物的分组或动态变量	
纵	一般为叙述事物的统计指标	
线条	顶、低线，或标目线或合计线	
数字	阿拉伯数字（不用文字）	
	小数位数一致、位次对齐	
	不能有空	
备注	必要时	
图例		放在图右上角或标题的上方

一般统计表的常见错误举例1

表 2-15 第三组病人各年存活及死亡情况 (原表)

年份	病例数	住院期 死亡数	急性期 死亡数	住院期 病死率(%)	急性期 病死率(%))
1964	17	8	7	47.1	41.2
1965	13	5	4	38.5	30.8
1966	15	7	6	46.7	40.0
1967	15	6	6	40.0	40.0
1968	12	4	4	33.3	33.3
合计	72	30	27	41.7	37.5

表 2-16 1964~1968 年急性心肌梗塞患者的病死率 (修改表)

年份	病例数	死亡例数		病死率 (%)	
		住院期	急性期	住院期	急性期
1964	17	8	7	47.1	41.2
1965	13	5	4	38.5	30.8
1966	15	7	6	46.7	40.0
1967	15	6	6	40.0	40.0
1968	12	4	4	33.3	33.3
合计	72	30	27	41.7	37.5

一般统计表的常见错误举例2

表X 用新训练方法184人中、远距离投篮情况分析（原表）

中距离投篮	远距离投篮		合 计
	达标	未达标	
达标	16	36	52
未达标	56	76	132
合 计	72	112	184

表X 184人新方法训练后的投篮情况 （修改表）

中距离	远距离		合 计
	达 标	未达标	
达 标	16	36	52
未达标	56	76	132
合 计	72	112	184

*10投6中以上为达标，低于6中为未达标

分组简单次数分布表例题

下面是**100**名学生在某项测验中的成绩分数，试将它们制成一个次数分布表。

76.0	77.5	82.0	80.5	81.0	85.5	71.0	80.5	92.5	77.0
88.0	81.0	76.5	67.0	83.0	84.0	84.0	62.0	79.0	72.0
89.0	78.0	78.0	80.0	78.5	76.5	75.0	79.5	86.0	81.5
75.0	84.0	90.0	80.0	86.0	84.5	68.5	71.0	86.0	81.5
79.5	80.5	73.0	93.0	83.0	72.0	68.0	71.0	87.0	78.0
66.0	83.0	87.0	82.5	79.5	80.0	82.0	81.0	86.5	83.5
71.5	83.0	91.0	96.0	75.5	89.0	87.5	69.0	74.0	70.0
77.5	75.0	79.0	79.0	80.5	74.5	77.0	82.5	72.5	73.5
73.5	76.0	88.5	85.0	89.5	78.5	76.0	74.0	98.0	73.0
94.0	79.0	80.0	75.5	83.5	82.0	65.0	74.5	80.0	70.0

分组简单次数分布表例题

步骤:

- (1) 求极差range
 $R=98-62=36$
- (2) 决定组数和组距
 $K=1.87(N-1)^{\frac{2}{5}}$
 $=11.75$
- (3) 列出组段 (组限)
- (4) 划记计数
- (5) 编制成表

分组区间	划记次数	组中值	次数 (f)
96-98		97	2
93-95		94	3
90-92		91	4
87-89		88	5
84-86		85	5
81-83		82	5
78-80		79	5
75-77		76	5
72-74		73	5
69-71		70	5
66-68		67	5
63-65		64	5
60-62		61	5
合计			100

[返回](#)

相对次数分布表例题

分组区间	组中值	次数（f）	相对次数	百分次数（%）
96-98	97	2	0.02	2
93-95	94	3	0.03	3
90-92	91	4	0.04	4
87-89	88	8	0.08	8
84-86	85	11	0.11	11
81-83	85	17	0.17	17
78-80	79	19	0.19	19
75-77	76	14	0.14	14
72-74	73	10	0.10	10
69-71	70	7	0.07	7
66-68	67	3	0.03	3
63-65	64	1	0.01	1
60-62	61	1	0.01	1
合计		100	1.00	100

[返回](#)

累积次数分布表例题

分组区间	组中值	次数 (f)	向上累积次数		向下累积次数	
			实际累积次数	相对累积次数	实际累积次数	相对累积次数
96-98	97	2	100	1.00	2	0.02
93-95	94	3	98	0.95	5	0.05
90-92	91	4	95	0.95	9	0.09
87-89	88	8	91	0.91	17	0.17
84-86	85	11	83	0.83	28	0.28
81-83	85	17	72	0.72	45	0.45
78-80	79	19	55	0.55	64	0.64
75-77	76	14	36	0.36	78	0.78
72-74	73	10	22	0.22	88	0.88
69-71	70	7	12	0.12	95	0.95
66-68	67	3	5	0.05	98	0.98
63-65	64	1	2	0.02	99	0.99
60-62	61	1	1	0.01	100	1.00
合计		100				

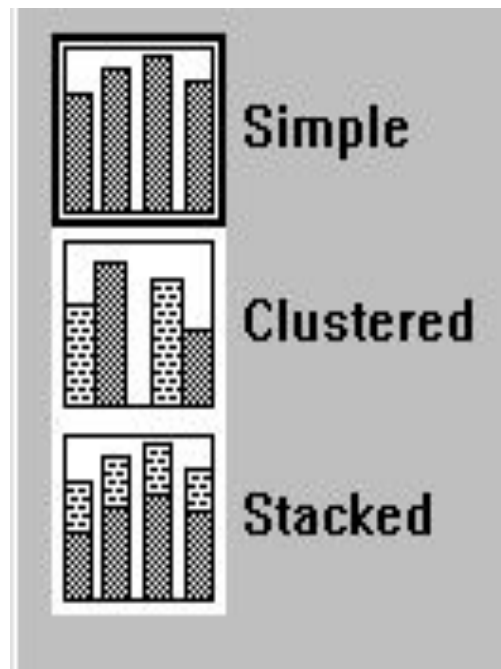
条形图 (bar chart)

条图：用直条的长度表示相互独立的统计指标的大小，可分为：

①单式 (simple) 条图：具有一个统计指标，一个分组标志。如图2-8。

②复式 (clustered) 条图：具有一个统计指标，两个分组因素。如图2-9。

③分段 (stacked) 条图



绘制条形图注意事项

(1) 纵轴的刻度**必须从“0”开始**，否则会改变各对比组间的比例关系。

(2) 横轴各直条一般按统计指标**由大到小排列**，也可按事物本身的**自然顺序排列**。

(3) 各直条的宽度要一致，各直条应有相等的间隔，其宽度一般与直条的宽度**相等**或为直条宽度的一半。

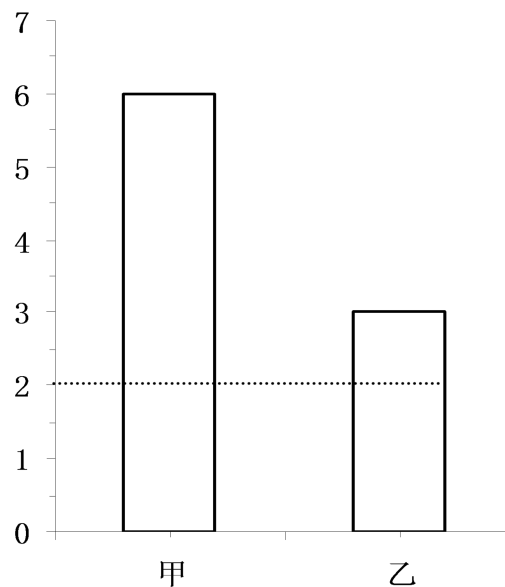


图2-10 直条图的纵轴尺度起点必须为零示意图

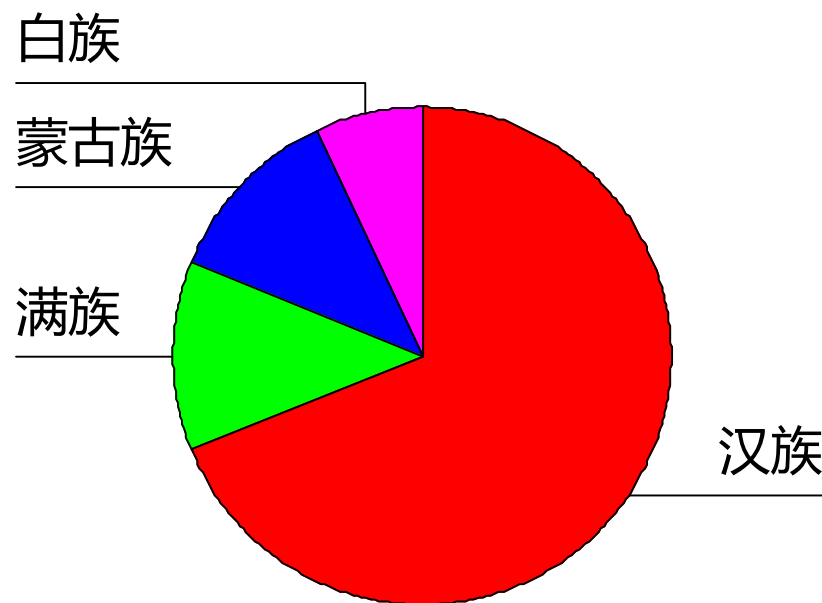
圆形图（pie chart）

圆图：用于表达各类事物的百分比构成。

➤各个扇形面积（**3.6°**）的大小反映了各组成部分百分比（**1%**）的大小。

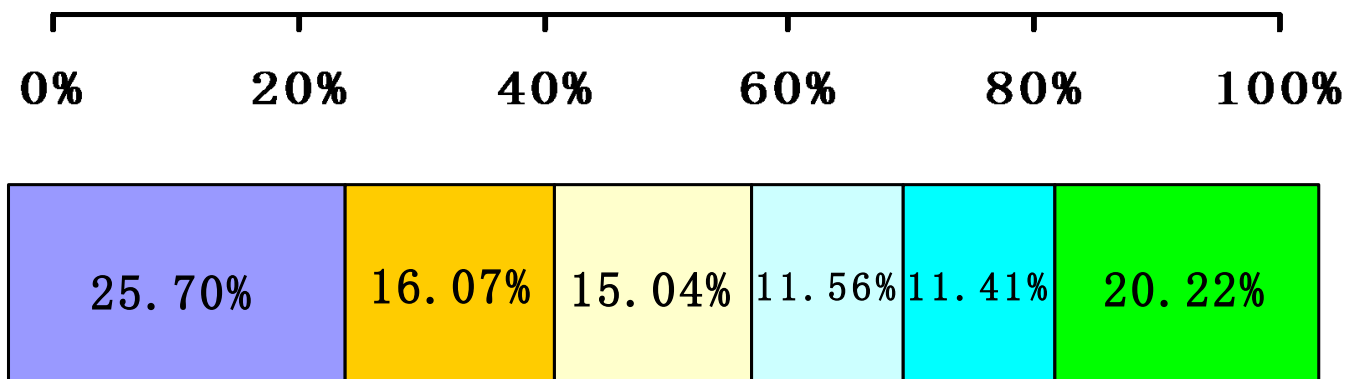
➤从**12点钟**处开始绘制，**顺时针**方向排列。

民 族	人数	构成比（%）
汉 族	69	69
满 族	12	12
蒙古族	12	12
白 族	7	7
总 计	100	100



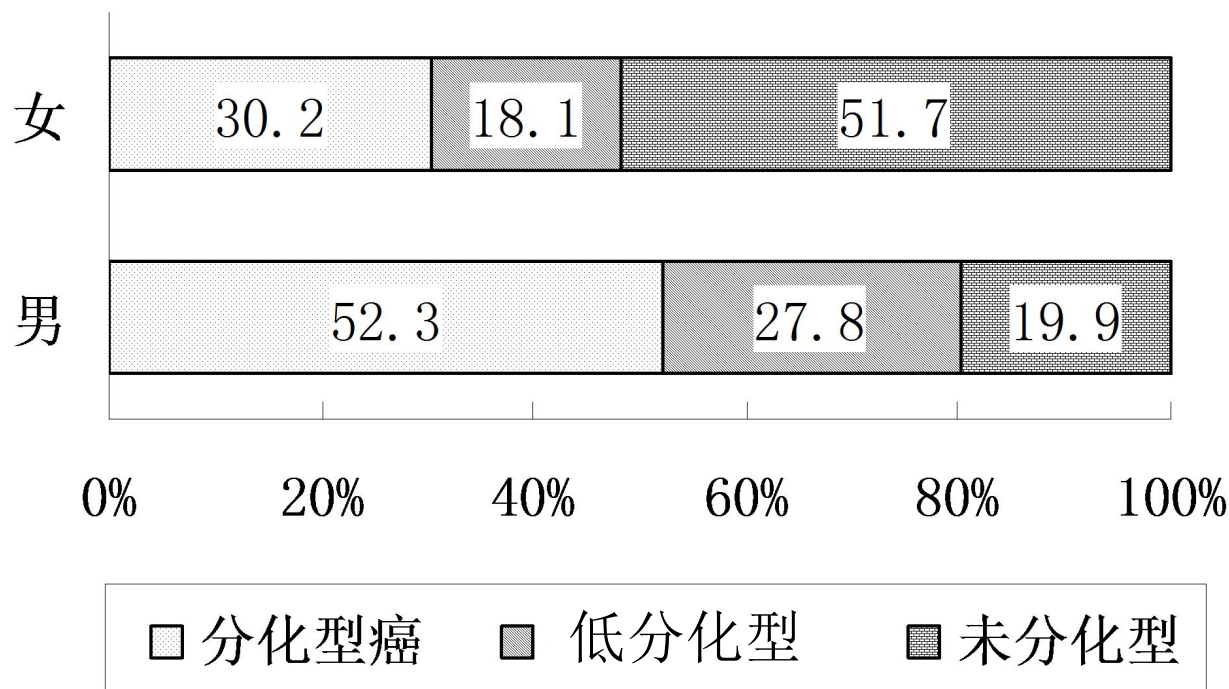
百分条图（percentage chart）

百分条图的作用与圆图相同。但更适用于多组百分比的比较。



不同性别某癌三种类型的构成

	分化型癌	低分化型癌	未分化型	合计
男	52.3	27.8	19.9	100
女	30.2	18.1	51.7	100



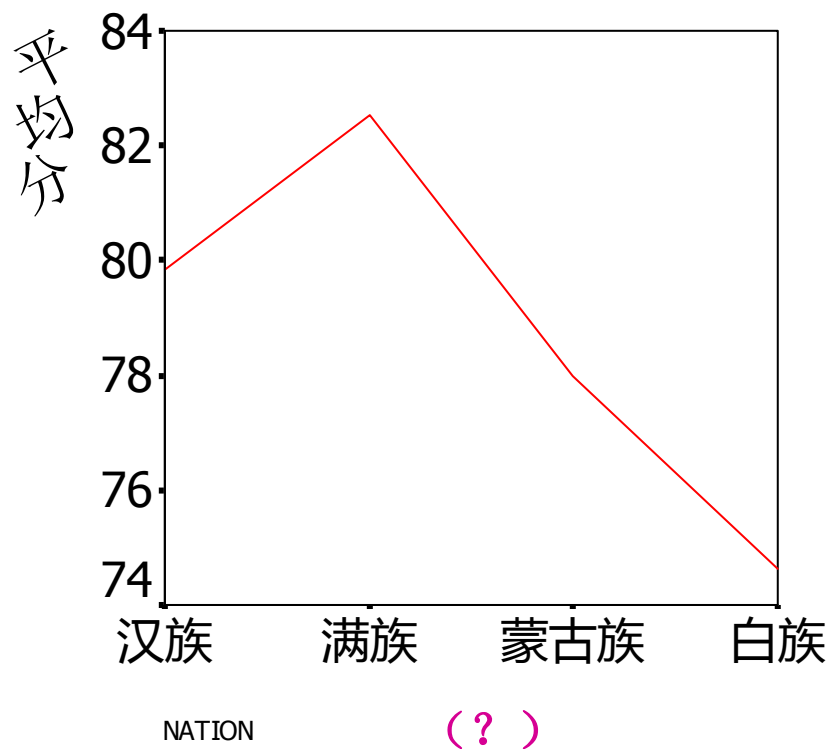
线图 (line chart)

线图(line chart): 用线段的升降表示某事物动态变化, 或某现象随另一现象变迁的情况。

适用于**连续型变量**。

纵轴: 算术尺度;

横轴: 连续型变量
(时间、年龄等)



直方图（histogram）

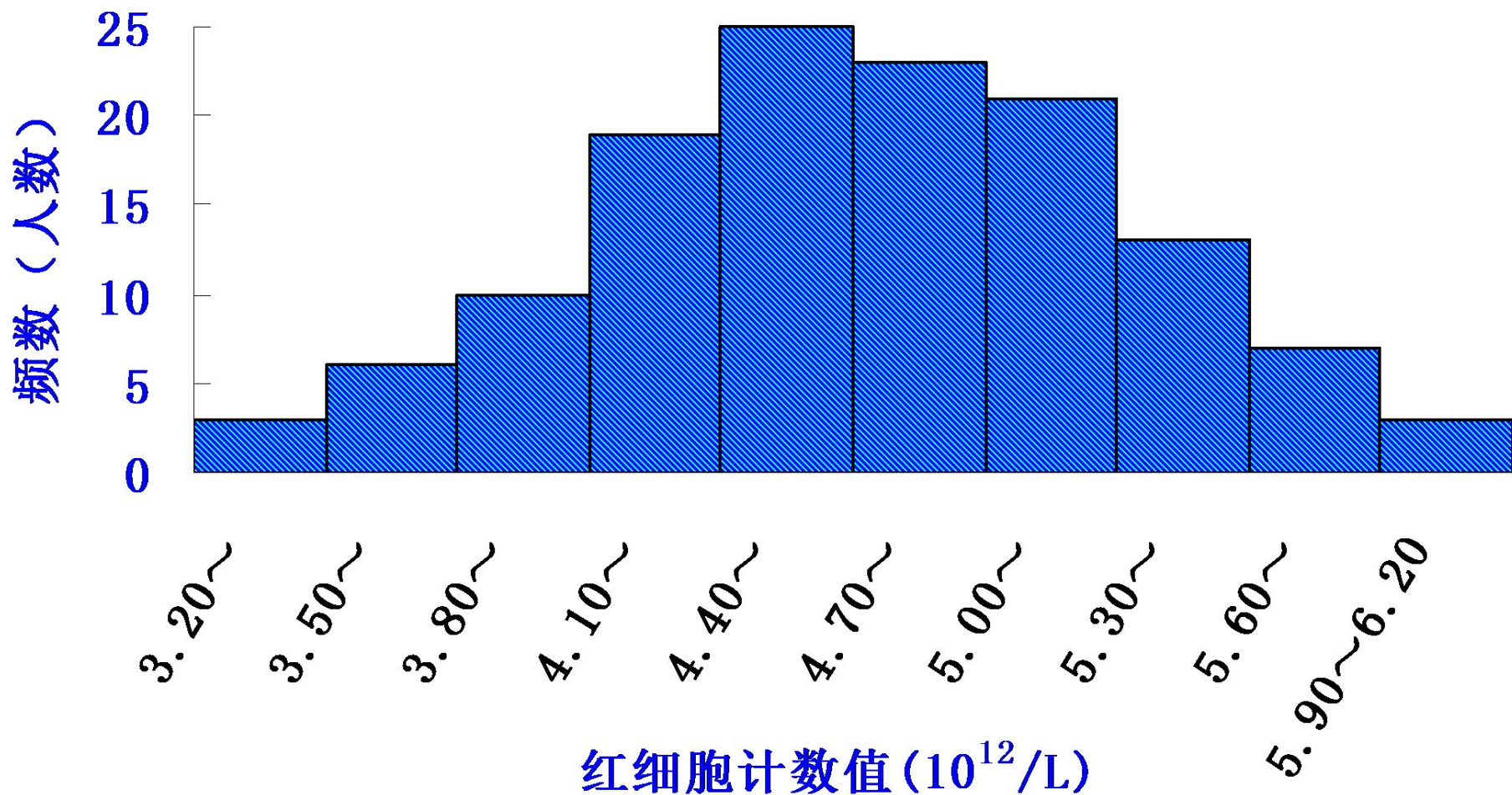
用矩形面积表示连续变量的次数分布。

1. 横轴：连续变量的组段（分组区间）；

纵轴：次数或次数密度（次数/组距）。

2. 各矩形条之间不留空隙。

3. 矩形的高度为次数或次数密度，宽度为组距。



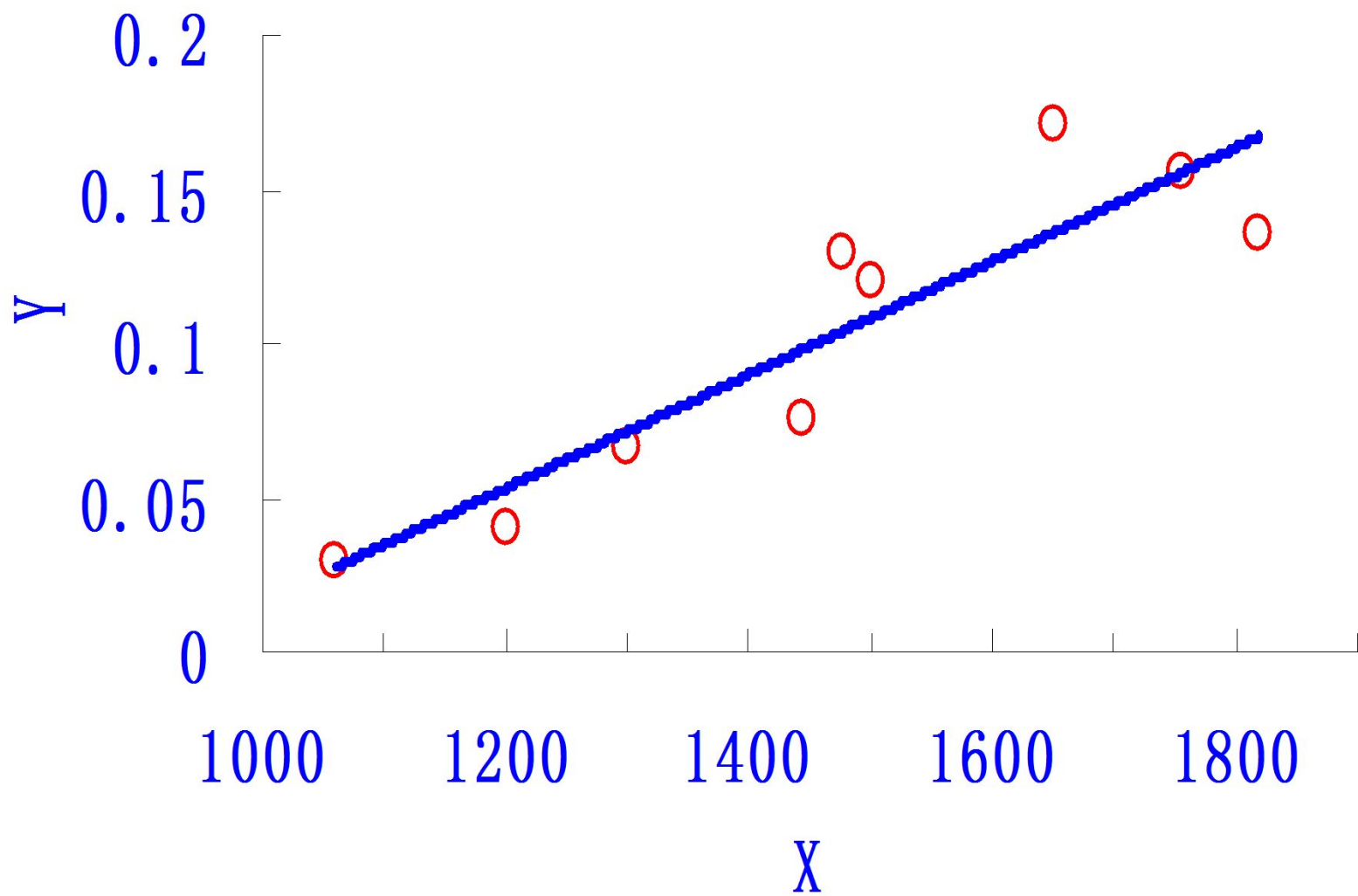
散点图(scatter diagram)

散点图(scatter diagram): 用点的密集程度和趋势表示两种现象或变量间的（相关）关系。

横轴：自变量 X

纵轴：应变量 Y

纵轴与横轴的起点可根据资料的情况而定。

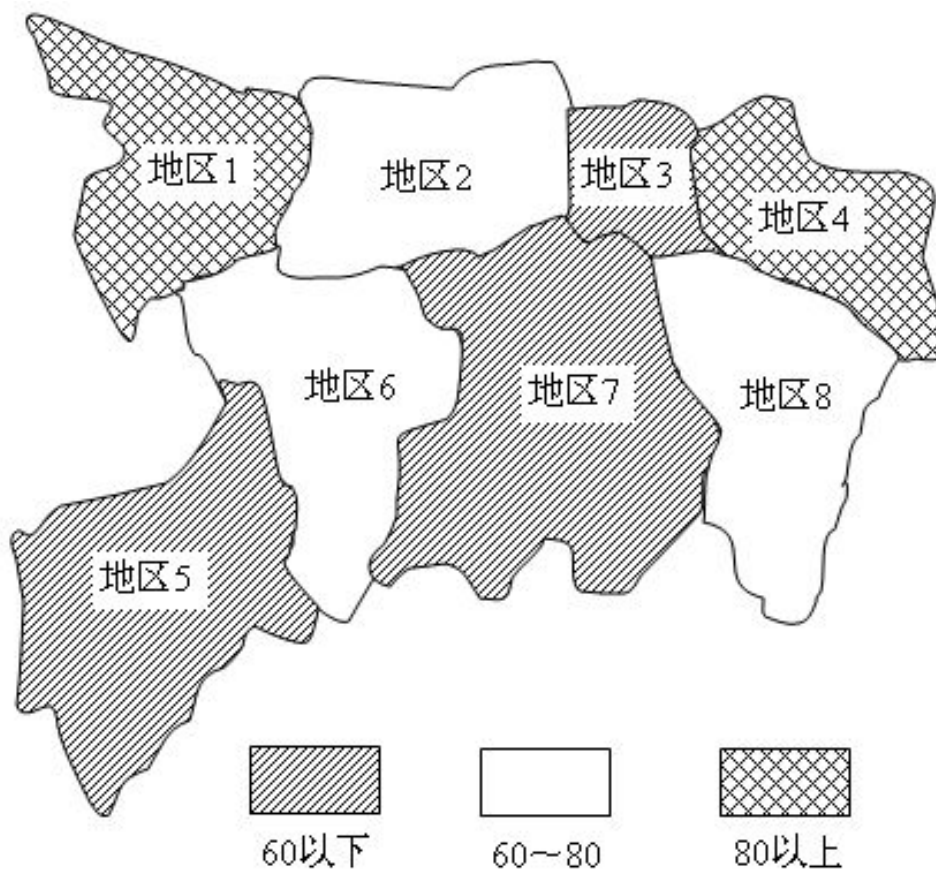


[返回](#)

统计地图(statistical map)

统计地图
(statistical map)

用于表示某现象
的**数量**在**地域**上
的分布。

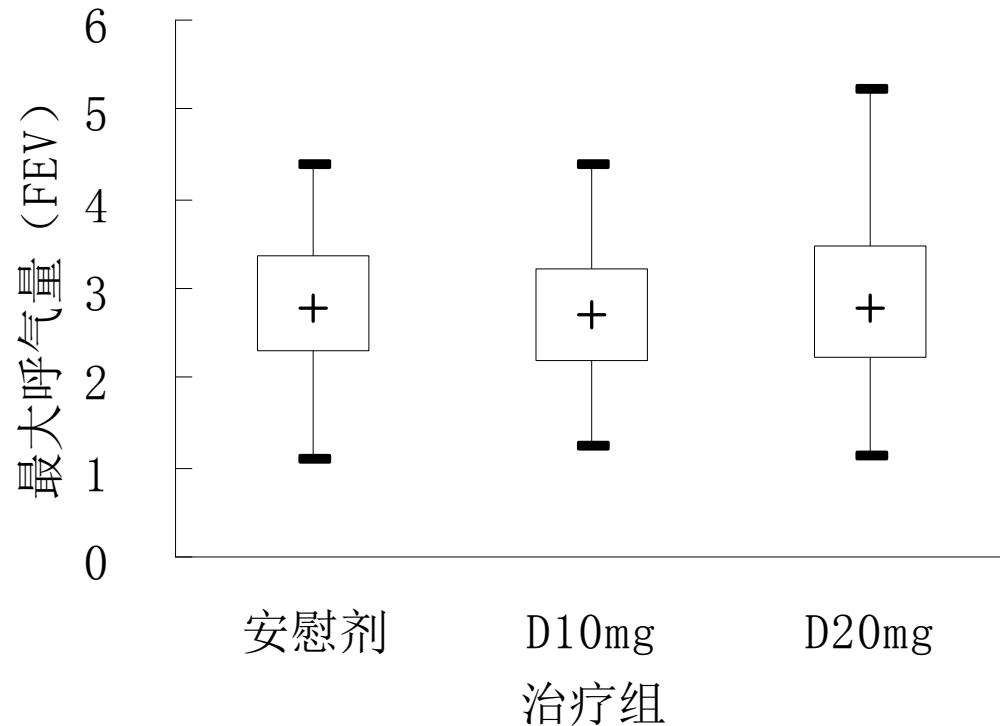


箱图(boxplot)

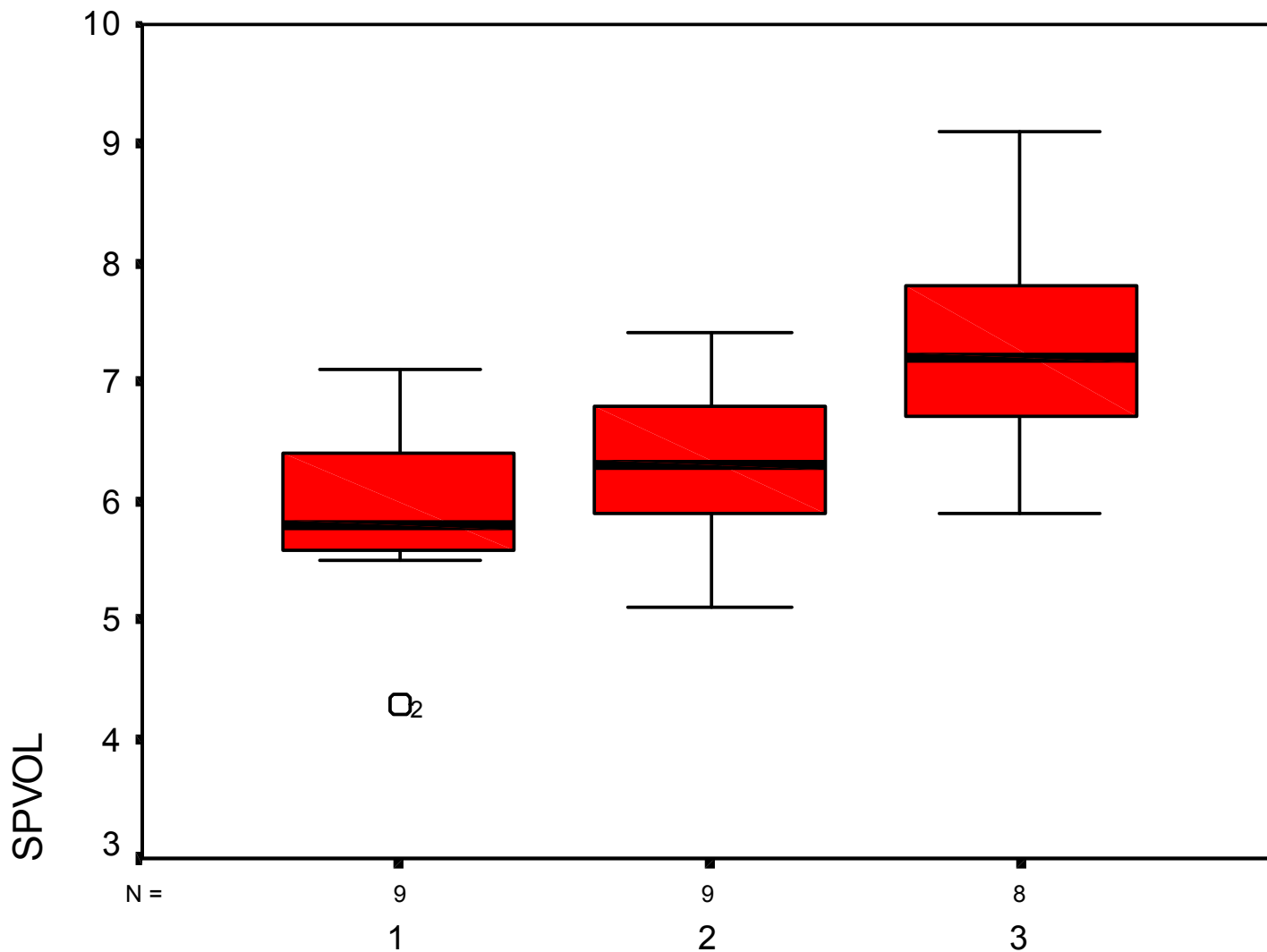
箱图 (boxplot)

也称箱须图

box-whisker plot,
用于反映一组或多
组连续型变量数据
的分布情况。



最小值 下四分位 中位数 上四分位 最大值



fat

如果所有样本中的数据都在离四分位点1.5倍盒子长度之内，则线的端点为最大和最小值.

[返回](#)

茎叶图(stem-and-leaf display)

茎叶图（ stem-and-leaf display ）与直方图类似，可描述连续性资料的分布，且直接用数据表示。

5	2	7														
6	2	2	2	2	5	6	7	8	8	8	9	9	9			
7	1	1	2	2	3	4	4	5	5	5	6	7	8	9	9	9
8	0	0	2	3	5	8	9									
9	1	3	7	7	7	8	9									
10	1	4	5	5	9											

HEIGHT Stem-and-Leaf Plot

<u>Frequency</u>	<u>Stem &</u>	<u>Leaf</u>
9.00	15 .	001223344
17.00	15 .	55666667778899999
20.00	16 .	0111222222333333444
35.00	16 .	5555566666666777778888888888999999
25.00	17 .	0000000011112222233333344
24.00	17 .	555666677777777777888899
13.00	18 .	0111111122333
11.00	18 .	55667788899
4.00	19 .	2333
2.00	19 .	56
3.00	Extremes	(>=198)

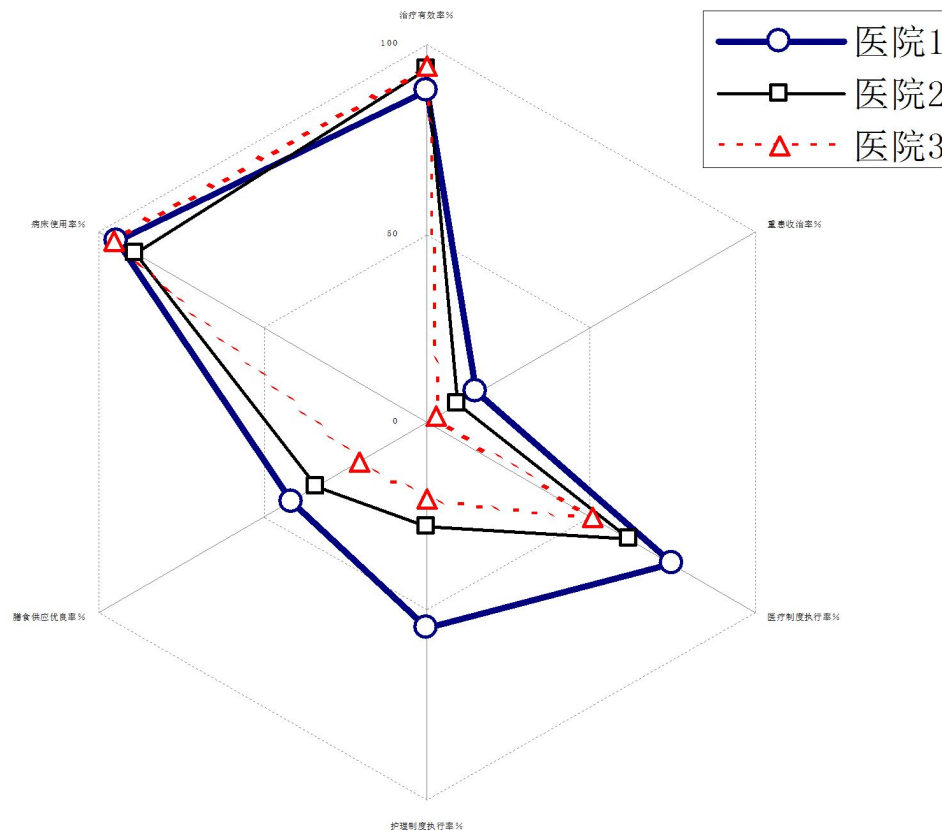
Stem width: 10.00

Each leaf: 1 case(s)

见茎叶图补充资料

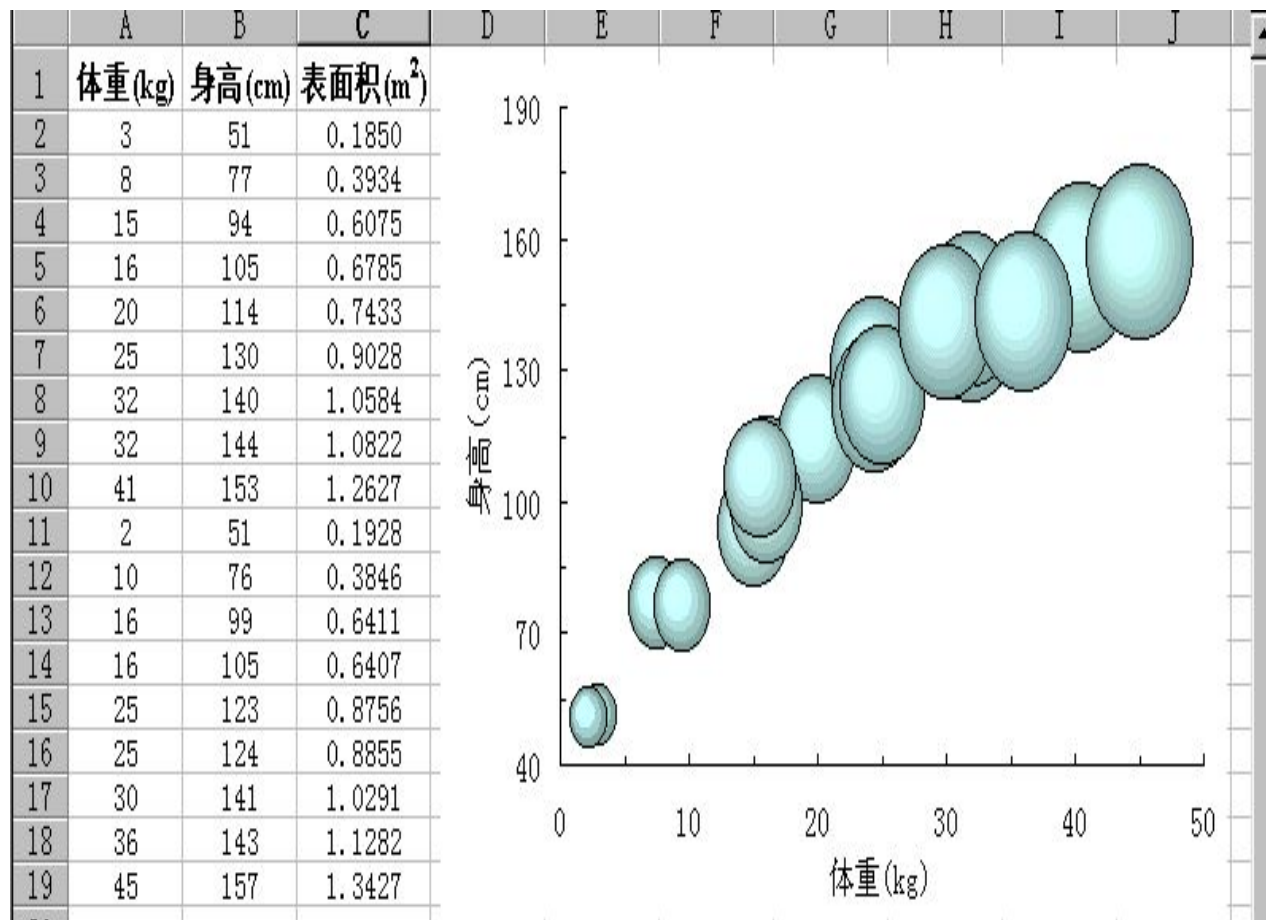
雷达图(radar graph)

雷达图（radar graph）可比较多个单位不同指标的数据特征，如最大值、最小值等。

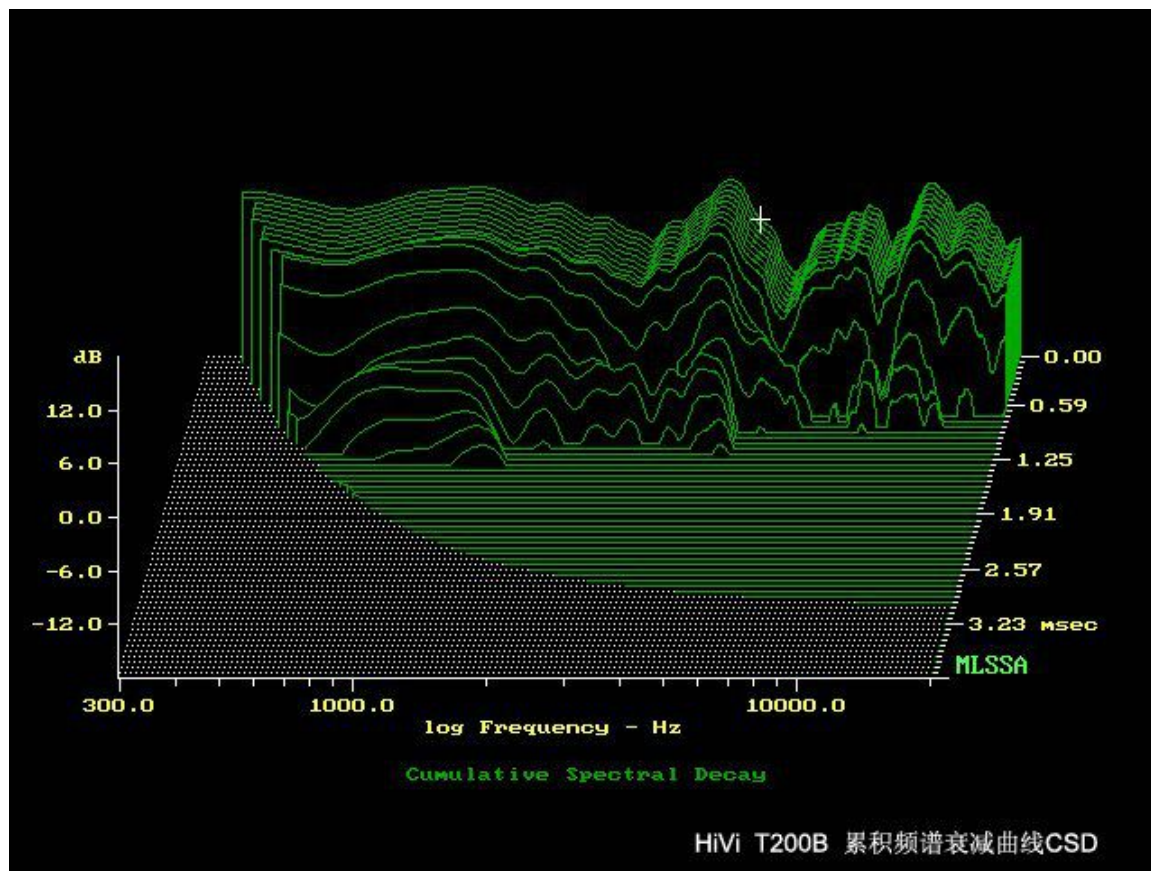


气泡图

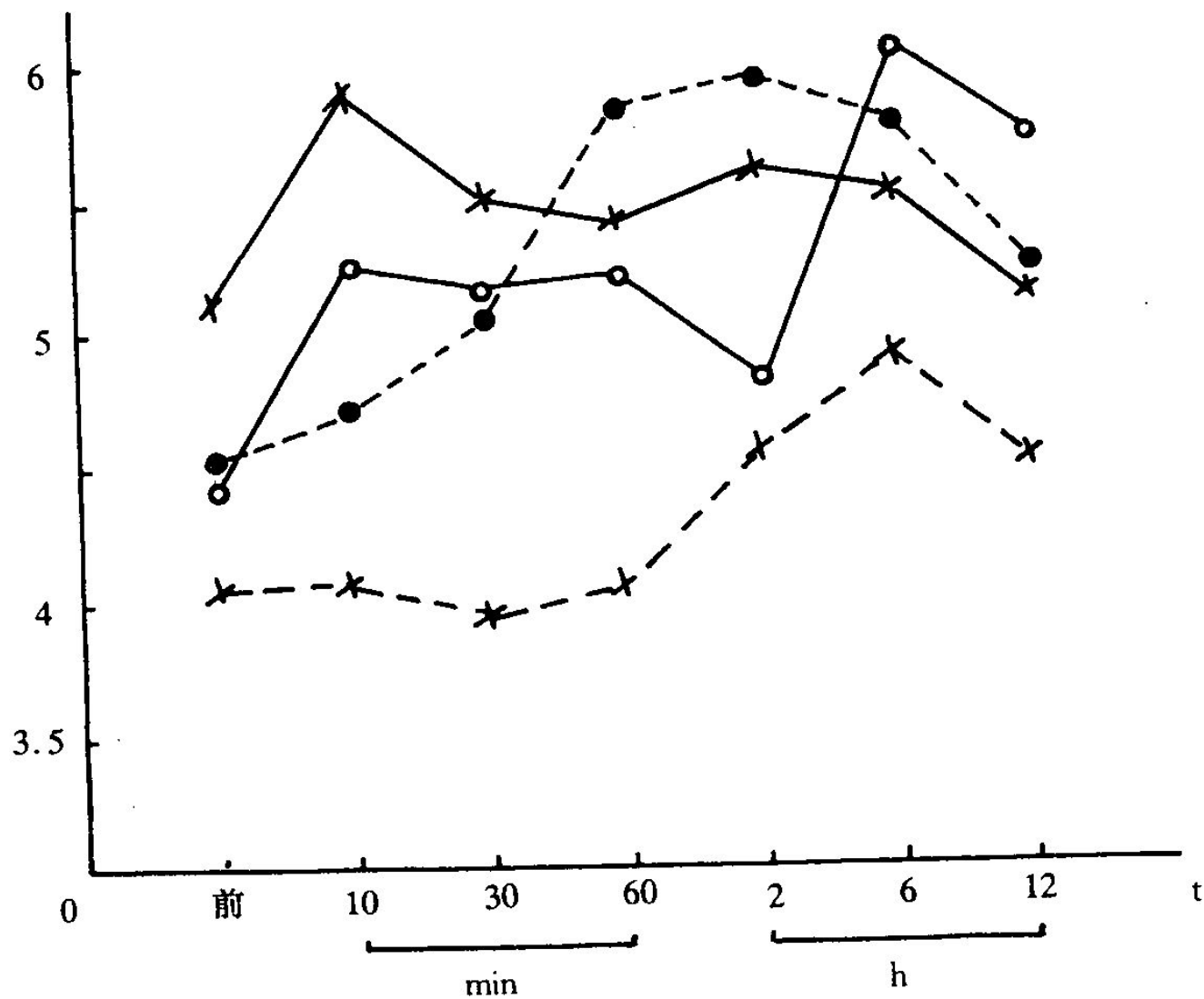
气泡图
可以反映三个
变量之间的数
量关系



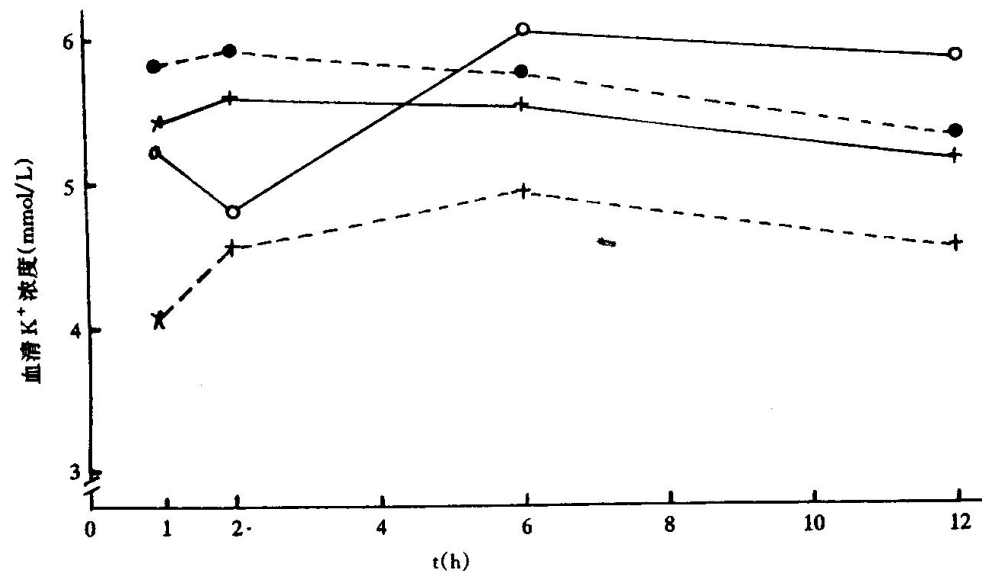
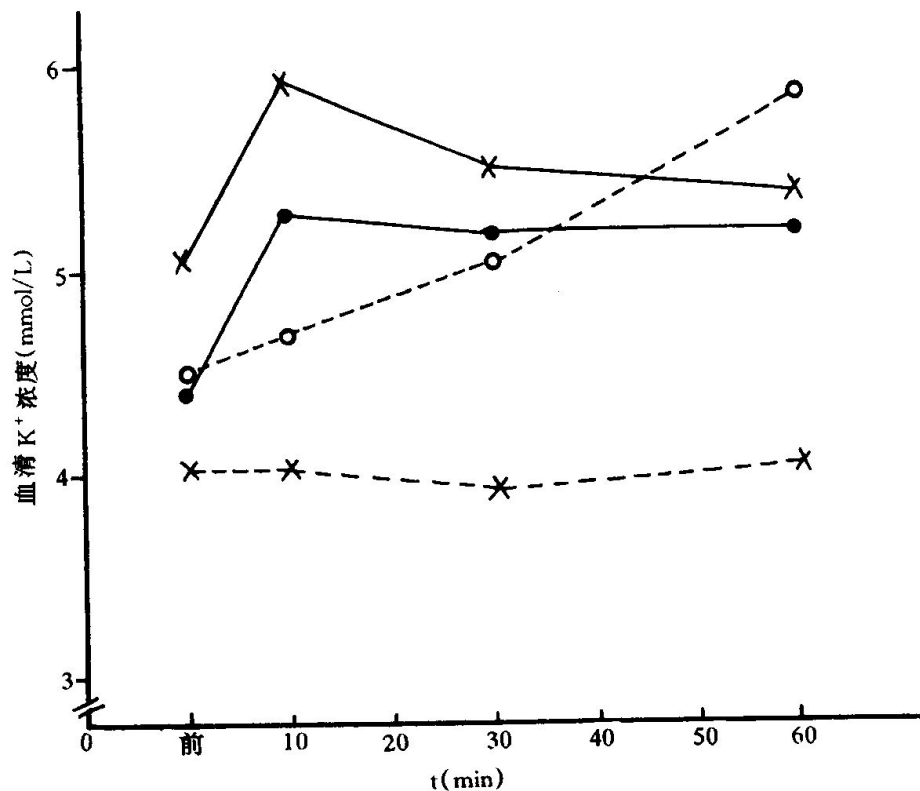
返回



图形法常见的错误例1



统计图中常见的错误 例1



修正图 1+1

[返回](#)

图形法常见的错误例2

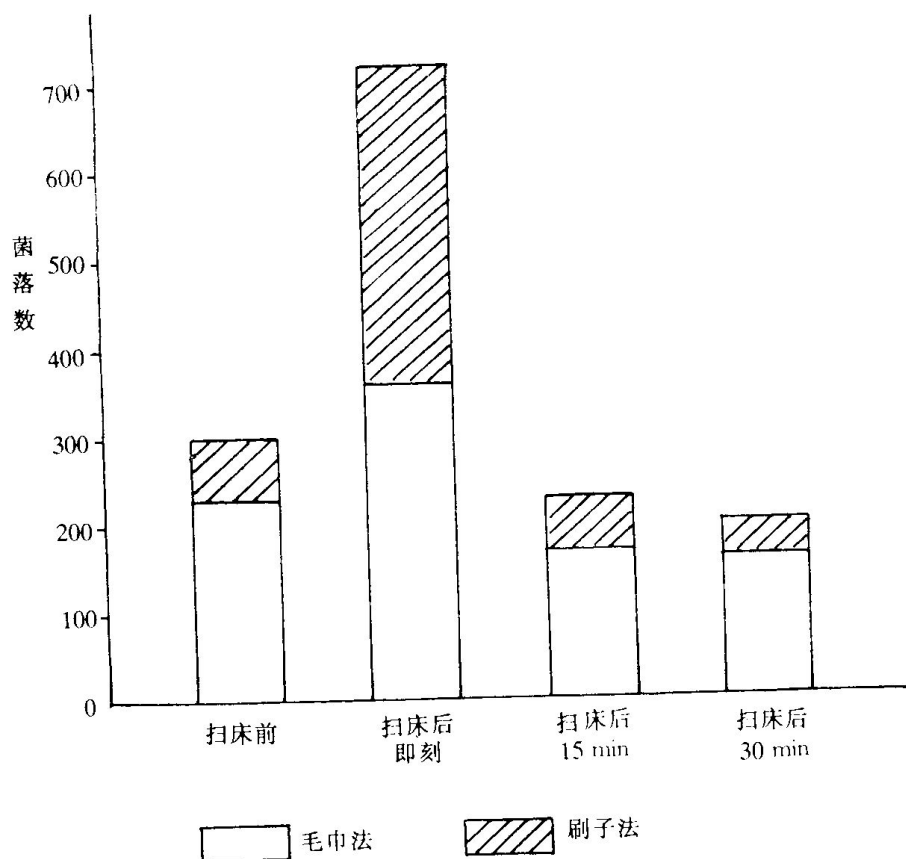


图 3-2(a) 两种方法扫床前后空气培养结果

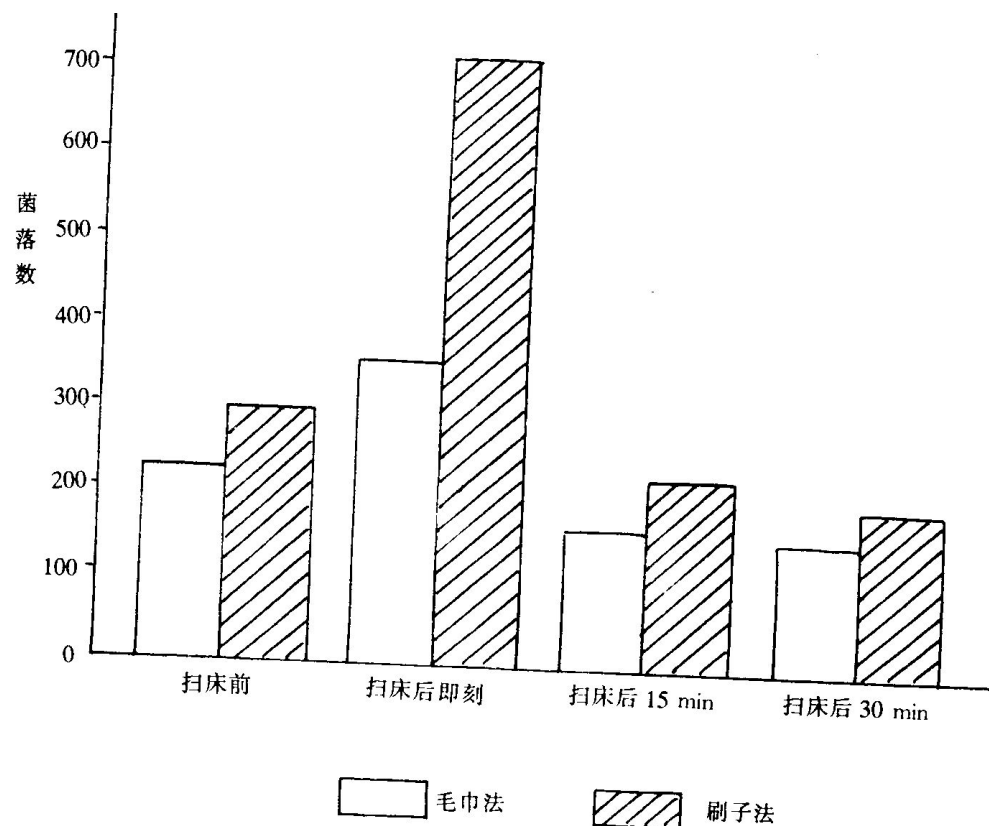


图 3-2(b) 两种方法扫床前后空气培养结果(修改后)

见教材46页

[返回](#)

再见！

