

# 第三章 社会统计资料的整顿

原始资料杂乱无章，需加整顿，才干为人所用。统计资料的整顿，其基础是统计分组。所谓统计分组，就是按统计研究的目的和要求，将总体单位或全部调查数据按一定的标志划提成若干组，使组内差别尽量小，而组与组之间则有明显差别，从而使原本杂乱无章的资料有序化，以便为在统计分析中提炼多种有用信息打下基础。

主要内容

统计分组的原则与原则

统计表

变量数列的编制

统计图



## 一、考核知识点

(一) 统计分组的原则与原则

(二) 统计表

(三) 变量数列的编制

(四) 统计图

## 二、考核要求

(一) 统计分组原则与原则

1、识记：(1) 统计分组；(2) 频数分配数列；  
(3) 品质数列和变量数列。

2、领略：统计分组的关键。

3、应用：在社会统计工作和研究中掌握统计分组的原则。

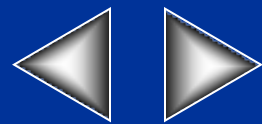


## (二) 统计表

- 1、识记：（1）统计表；（2）主词与宾词；（3）统计表格式；
- 2、领略：简朴表、简朴分组表、复合分组表和简朴设计、复合设计。
- 3、应用：掌握统计表制作的一般规则。

## (三) 变量数列的编制

- 1、识记：（1）变量；（2）频数；（3）相对频数（频率）；（4）单项式变量数列和组距式变量数列；（5）等距与异距；（6）开口与闭口；（7）上限、下限和组距；（8）组中值；（9）频率（频数）密度；（10）合计频数（合计频率）。
- 2、领略：组数、组距和全距三者的关系。
- 3、应用：（1）对离散变量编制单项式变量数列或组距式变量数列；（2）对连续变量编制组距式变量数列；（3）掌握频数（频率）密度和原则组距频数（频率）的换算措施；（4）掌握向上（或向下）合计频数（或频率）。



## （四）统计图

- 1、识记：（1）直方图；（2）折线图；（3）曲线图；（4）合计频数（频率）分配曲线；（5）正态分布和偏态分布；（6）洛仑兹曲线。
- 2、领略：（1）多种统计图之间关系；（2）洛仑兹曲线和基尼系数的意义
- 3、应用：（1）学会绘制直方图、折线图、曲线图及合计频数（频率）分配曲线；（2）学会绘制洛仑兹曲线；（3）学会计算基尼系数。



# 第一节 统计分组的原则与原则

## 统计分组的原则

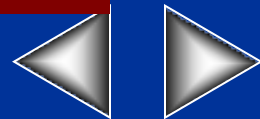
### 1. 分组应使各类别构成之和等于总体

“穷举”

“互斥”

### 2. 分组设计应能反应统计总体的分布规律性

统计分组主要是为了能很好地反应统计总体的构成情况，即反应总体中各单位的分布特征。分组设计要适应这一要求，必须在分组后使总体单位总数在各组的分配情况能够反应总体的分布规律性。



# 频数分布数列

变量数列

品质数列

在统计资料搜集的基础上，按分组原则，将总体中全部单位依一定顺序归类整顿，即可得到能够表白总体单位总数在各组分配情况的频数(或次数)分布数列，简称数列。频数分布数列是统计分组工作的产物。

## 某社团组员的构成

性别	人数	年龄	人数
男	28	15--25	3
		25--35	7
女	22	35--45	12
		45--55	18
		55--65	10

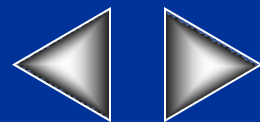


## 第二节 统计表

100名男性青年的身高表 单位：厘米

166	181	173	171	168	164	178	167	156	194
169	175	183	168	165	180	170	172	167	175
173	169	174	172	171	169	173	184	173	169
170	181	185	168	174	187	167	174	169	165
167	172	174	173	155	179	164	158	165	162
157	175	182	168	176	161	171	163	176	159
171	178	172	169	158	163	166	168	160	178
170	164	160	170	166	178	171	167	162	169
165	171	165	168	176	174	163	177	164	170
161	179	177	162	149	169	166	153	177	164

统计调查搜集来的资料往往是没有顺序的原始资料(参见右表),使原始资料有序化,列表和作图是两种基本措施,得到的分别就是统计表和统计图。



# 1. 统计表的格式、内容与种类

## ●格式

统计表是表达统计资料的表格，在由横行、纵栏交叉结合而成的表格上，它能系统地组织和合理地安排大量数字资料。

总标题

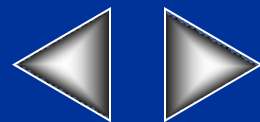
纵栏标题

2023年我国城乡人口情况

按城乡分	人口数（万人）	比重（%）
城乡	50212	39.09
乡村	78241	60.91
合计	128453	100.00

横行标题

统计数值





## ●内容

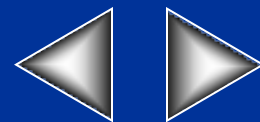
主词——统计表所要阐明的对象。

宾词——用来阐明主词的标志和标志值。

如需要，主宾词可互换。

表3. 4 离婚案件构成

	草率型	外遇型	猜疑型	虐待型	再婚型	家务型	生理型	分居型	其他型	合计
离婚案件 (件)	85	79	64	44	44	188	46	14	59	623
比重 (%)	14	13	10	7	7	31	7	2	9	100



●统计表按主词是否分组以及分组的程度可分为

简朴表

简朴分组表

复合分组表

■ ●简朴表 我国城市居民家庭基本情况

(2023年)主要指标	单位	数值
■ 平均每户家庭人口	人	3. 04
■ 平均每户就业人口	人	1. 58
■ 平均每个就业者承担数	人	1. 92
■ 平均每人全部年收入	元	8177. 40
■ 平均每人可支配收入	元	7702. 80
■ 平均每人消费性支出	元	6029. 88



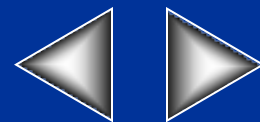
- 简朴分组表

中国人口年龄构造情况

单位：%

年龄组	1983年	1994年	■ 2023年	■ 2023年	■ 2023年
0—14岁	36.3	40.7	33.6	27.7	19.9
15—64岁	59.3	55.7	61.5	66.7	70.1
65岁以上	4.4	3.6	4.9	5.6	10.0

资料起源：《2023中国统计年鉴》第99页。



## ● 复合分组表

### 我国社会福利主要费用情况

单位：亿元

项目	■ 2023	■ 2023	■ 2023	■ 2023
■ (一) 优抚对象补贴金额	68.0	91.4	107.6	
■ 国家支出	32.4	51.1	60.7	
■ 集体供给	35.6	40.3	46.9	
■ (二) 农村老式救济金额	29.8	28.5	31.7	
■ 国家支出	7.0	7.9	8.3	
■ 集体供给	22.8	20.6	23.4	108.1
■ (三) 城乡多种福利金额	20.2	23.1	28.7	69.5
■ 国家支出	10.3	13.7	19.0	38.6
■ 集体供给	9.9	9.4	9.7	29.5
■ (四) 城乡多种福利				12.1
■ 国家支出				17.4
■ 集体供给				40.5
■ (五) 城乡多种福利				26.4
■ 国家支出				14.1
■ 集体供给				

资料来源：《2023中国统计年鉴》第838页。

## 2. 统计表的制作规则

### ✓ 标题部分

- 1、标题置于表格正上方
- 2、总标题所示要点与表中项目一致，在需要时还应表白资料所属的时间和地域
- 3、表次：左；单位：右
- 4、对分页的同一表格，在每页上端都要写标题，加（续一）、（续二）

### ✓ 栏目部分

- 1、先局部后整体
- 2、若栏目较多，可加以编号；统计数字间有计算关系的，可用计算式体现。



## ✓ 线格部分

1、统计表上下两端线应以粗线或双细线标划，表的

左右两侧开口。

2、各栏间用直线标划，大项目间线条较粗，小栏目

线条较细；各行间不必划线条。

## ✓ 数字部分

1、表中数字要对准位数。

2、不存在某数字时，用“——”表达；缺乏某项数字

时，用“……”表达。

3、数字较大时，加分位点。

## ✓ 其他规则

1、资料起源写在表格下方。

2、有阐明解释需要时，在表下方注释。

3、单位有数种时而不能在表右上角划一标注时，

分

1. 统计表的纵栏标题是用来阐明 ( )

A统计表的名称

B各组的名称

C统计指标的名称

D统计总体

2. 填统计表时, 若某项数据缺失, 其符号为 ( )

3 统计分组的成果体现为 ( )

A组内同质性组间差别性

B 组内差别性组间同质性

C组内同质性组间同质性

D组内差别性组间差别性



4 下面属于按数量标志分组的是 ( )

A 企业按经济类型分组

B 工人按民族分组

C 学生按考试成绩分组

D 人口按性别分组

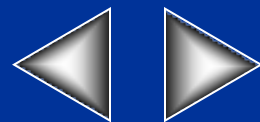
5 统计分组的关键是 ( )

A 按数量标志分组

B 正确选择分组标志

C 正确计算组距和组中值

D 利用统计体系分组





## 第三节 变量数列的编制

### 变量数列的构成要素

在社会统计学中，总体中各单位的分布特征首先是用统计表来表达的。能够表达变量分布及其特征的统计表，即变量数列。它有两个构成要素：

变量值  $X_i$

频数  $f_i$

相对频数  $P_i$

$$P_i = \frac{f_i}{N}$$

$$\sum_{i=1}^n f_i = N$$

$$\sum_{i=1}^n P_i = 1$$



# 1. 离散变量数列

离散变量所描述的对象的数量特征，能够按一定顺序列出它的整数值，相邻两变量值不会出现小数，因而能编制出单项式和组距式两种变量数列。

单项数列

组距数列

组距( $h_i$ ) = 上限( $u_i$ ) - 下限( $l_i$ )

$$\text{组中值}(m_i) = \frac{l_i + u_i}{2} \quad \text{或} \quad \text{组中值}(m_i) = l_i + \frac{u_i - l_i}{2}$$

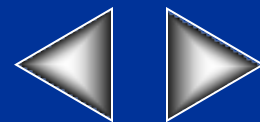


## • 离散变量单项数列

单项式变量数列——数列中每一组的变量值只有一个。单项数列用于离散变量整数数值变动幅度较小时。

某小区各户人口数统计表

人口数 (X)	户数 (f)	频率 (P)
2	5	0.10
3	8	0.16
4	16	0.32
5	10	0.20
6	6	0.12
7	4	0.08
8	1	0.02
合计	50	1.00



## ● 离散变量组距数列

组距式变量数列——离散变量的整数值假如变动幅

度较大，而且总体单位数N又很大，则要编制组距数列。  
某地域小学班级数

间距	组中值 ( $X_i$ )	学校数 ( $f_i$ )
8—14	11	2
15—21	18	5
22—28	25	10
29—35	32	8
36—42	39	4
合计	—	29



## 2. 连续变量数列

连续变量因其数学特征，在一种区间能够有无限多数值，无法按顺序一一列举，所以只能编制组距数列。与离散变量组距数列不同之处在于，根据连续变量的特征，此时组距数列中相邻两组的上限和下限共有一种组限，即相邻两组交界处的组限重叠。至于恰等于某一组限的数据(如下表中身高164厘米)归于哪一组，应该按照“上限不涉及在内”的原则处理。这就是说，164应归于“164—168”这一组，而不应归于“160—164”这一组。有了这一要求，就不会在编制连续变量的数列时，发生违反“穷举”与“互斥”这两个基本原则的情况了。



## ● 连续变量数列

男青年身高按4厘米的间距分组时的频数分布

身高间距（厘米）	组中值 ( $X_i$ )	频数 ( $f_i$ )	频率 ( $P_i$ )
148—152	150	1	0.01
152—156	154	2	0.02
156—160	158	5	0.05
160—164	162	10	0.10
164—168	166	19	0.19
168—172	170	25	0.25
172—176	174	17	0.17
176—180	178	12	0.12
180—184	182	5	0.05
184—188	186	3	0.03
188—192	190	0	0.00
192—196	194	1	0.01
合计	——	100	1.00



### 3. 拟定组距和组数

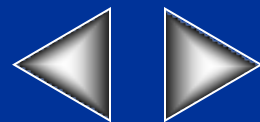
对于等距分组且为闭口组的情况，拟定组距已经有某些数学公式可供参照

$$h = \frac{R}{1 + 3.322 \lg N}$$

但最佳决定还是根据常识和数列使用的目的而定。一般地说，组距应不不大于能够忽视的数值之差。

$$\text{组数 } (n) = \frac{\text{全距 } (R)}{\text{组距 } (h)}$$

注意，在资料被整顿成数列时，全距可合适放大  
(但不能缩小)，以便组数(或组距)取整数值。



## ● 异距分组

异距分组主要在变量变动并不是均匀的、有急剧上

升或忽然下降之类情况发生时考虑。

- $\text{频数密度} = \frac{\text{频数}}{\text{组距}}$

**原则组距频数**的换算措施：

- (1) 选定某一合适的组距为原则组距；
- (2) 用原则组距除以各组组距，得到折合系数；
- (3) 将各组的折合系数乘以各组的频数。

- $\text{频率密度} = \frac{\text{频率}}{\text{组距}}$





## 频数密度和原则组距频数计算表

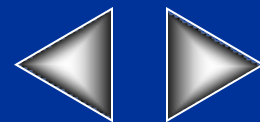
男青年按身高分组（厘米）	频数	组距	频数密度	■ 折合系数 ■ （取原则组距4）	■ 原则组距频数
148—156	3	8	$3/8$	$1/2$	1.5
156—164	15	8	$15/8$	$1/2$	7.5
164—168	19	4	$19/4$	1	19.0
168—172	25	4	$25/4$	1	25.0
172—176	17	4	$17/4$	1	17.0
176—180	12	4	$12/4$	1	12.0
180—188	8	8	$8/8$	$1/2$	4.0
188—196	1	8	$1/8$	$1/2$	0.5
合计	100	——	——	—	

## 4. 合计频数 (F)

■ 向上合计——以变量数列首组的频数为始点，逐个合计各组的频数，展示不大于该组上限的频数和。

■ 向下合计——以变量数列末组的频数为始点，逐个合计各组的频数，展示不小于该组下限的频数和。

组距	f	F 向上累计	F 向下累计
148~156	3	3	100
156~162	9	12	97
162~168	25	37	88
168~174	34	71	63
174~180	20	91	29
180~186	7	98	9
186~192	1	99	2
192~198	1	100	1
合计	100	—	—



1. 连续变量要求两个相邻组的组限 ( )

A 不一致

B 重叠

C 间断

D 可取任意值

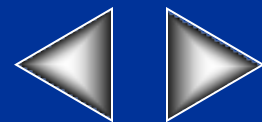
2. 在全距一定的条件下, 等距分组中组距与组数的关系是 ( )

A 组数越多, 组距越大

B 组数越多, 组距越小

C 组数越小, 组距越小

D 组数与组距无关系



3 在分组时，但凡遇到某个体的变量值恰好等于相邻两组上下限数值时，一般是（ ）

A 将此值归入上限所在组

B 将此值归入下限所在组

C 将此值归入上，下限所在组均可

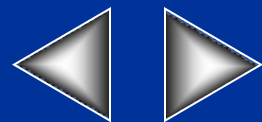
D 另立一组

4. 对离散变量进行统计分组（ ）

A 只能用单项分组

B 既可用单项分组也可用组距分组

C 只能用组距分组



# 第四节 统计图

频数分布不但能够用统计表的形式体现，也能够用

统计图的形式体现。用统计图表达频数分布，较之用统计表，要直观便捷得多。但缺陷是不及统计表精确。统计图的种类诸多，本节仅就与频数分布数列相衔接的统计图加以简介。

根据编制好的频数分布数列，能够绘制出相应的统

计图，最常用的有频数分布直方图、折线图、曲线图以及合计频数分布曲线。详细措施是：先画直角坐标系，横轴代表分组或各组组限，纵轴代表各组频数或频率，然后再根据相应的分配数列作图。



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/707115123143006141>