

项目三

统计整理

学习目标

本项目主要介绍统计数据的整理过程。通过本项目的学习，掌握统计整理的概念和意义；了解统计分组在整个统计工作中的重要作用，以及统计整理的基本步骤和方法；掌握选择分组标志、编制变量数列、制作基本统计图表的技术和方法。

任务一 统计整理概述

统计资料整理是统计工作的第三个阶段，在整个统计工作过程中起着承前启后的作用。统计整理既是统计调查的继续和深化，又是统计分析的基础和前提，是统计调查和统计分析的连接点。

一、统计整理的概念和意义

统计资料整理简称统计整理，是指根据统计研究的目的和任务，对统计调查所得的原始资料进行科学的分类和汇总，或对已初步加工的次级资料进行再加工，使其系统化、条理化、科学化，以反映所研究的现象总体特征的工作过程。

为什么要对数据进行统计整理？统计调查得到的大量、零散、不规范的资料只能表明被调查单位的具体情况，反映事物的表面现象，不能说明事物的综合情况，必须进行加工整理。

例如，人口普查中的人口资料只能说明每一个人的具体情况，如姓名、性别、年龄、文化程度等。必须通过对人口总体中每个人的资料进行整理、分组、汇总等加工处理后，才能得到人口总体的综合情况，如总人口数、平均年龄、男女性别比例等，从而了解人口总体的规模、结构、增减变动状况等，达到对人口总体的全面、系统的认识。

因此，统计整理是实现由对个别现象到总体现象的认识，统计整理的正确与否、质量好坏，将直接影响统计对社会经济现象数量描述的准确性和数量分析的真实性。采用科学

的方法进行统计整理是顺利完成统计分析任务的前提。

二、统计整理的内容

统计整理的全过程包括对统计资料的审核、统计分组、统计汇总和编制统计图表4个环节。

▶ 1. 统计资料的审核

原始资料一经汇总，资料中的各种差错就会被掩盖起来，会影响统计资料的质量，对调查资料进行审核是统计整理的第一步，包括以下内容：①审核资料的完整性，即调查单位或填报单位是否齐全；②审核资料的及时性，即填报单位是否按规定的时间报送了有关资料；③审核资料的正确性，即检查所填报的资料是否准确、可靠。

资料正确性的审核较为复杂，我们常采用以下两种方法。

(1) 逻辑性审核。审核资料的回答内容是否合情合理，各个项目之间是否有矛盾。例如，在某省人口调查表中，“年龄”一栏中填写“10”，而“从事工作”一栏中却填写“财务”，这其中必有一栏填写错误。

(2) 计算审核。审核资料的统计口径、范围、计算方法和计量单位是否符合要求，统计数字是否正确。

▶ 2. 统计分组

根据研究目的和统计分析的需要，选择整理的标志，并进行归类分组。统计分组是统计整理的重要内容和统计分析的基础，只有正确的分组才能整理出有科学价值的综合指标，并借助这些指标来揭示现象的本质与规律。

▶ 3. 统计汇总

统计汇总就是对分组后的各项统计指标进行汇总处理，并计算各组的单位数和合计数，计算出说明总体和各组情况的统计指标数值。

▶ 4. 编制统计图表

通过编制统计表和绘制统计图，将整理出的资料简洁、明了且有序地呈现在读者面前。

任务二 统计分组

一、统计分组的相关概念

统计分组是根据统计研究的目的和任务，按照选定的变异标志将总体划分为若干部分或组别，使组与组之间具有差异性，而同一组内的单位保持相对同质性。例如，社会产品按其经济用途分为生产资料和消费资料；企业按年产量或投资总额可分为大型企业、中型企业、小型企业等。通过统计分组，可以区别社会经济现象在质和量方面的差别。

统计分组在统计研究中的作用主要体现在以下三个方面。

► 1. 区分社会经济现象的类型

这类分组称为类型分组。社会经济现象是极其复杂多样的，客观上存在各种不同的类型，各种不同类型的现象在规模、水平、结构、比例关系等方面的数量表现也各不相同。利用统计分组就能根据统计研究的目的，将总体区分为各种性质不同的类型，来研究各类现象的数量差异和特征以及相互关系。例如，中国居民城镇、农村的划分就说明在中国二元经济结构下社会经济关系存在巨大的差异。

► 2. 研究总体内部结构

这类分组称为结构分组。利用统计分组计算出各组数值在总体中所占比重，对社会经济现象的内部结构进行研究，可以说明现象总体的基本性质和特征。同时，对现象内部结构的变化进行动态研究，还可以反映现象总体发展变化的过程、趋势和规律。

【例 3-1】某高校的经济管理学院近几年教师学历结构如表 3-1 所示。

表 3-1 某高校经济管理学院近几年教师学历结构表

年份	本科/%	硕士/%	博士/%	合计/%
2013	50	35	15	100
2014	44	38	18	100
2015	40	36	24	100
2016	38	36	26	100
2017	34	38	28	100

从表 3-1 中可以看出，该学院通过近几年的教师队伍学历优化，初步实现了提高教师队伍学历的目标。

► 3. 研究现象之间的依存关系

这类分组称为分析分组。客观现象之间都存在不同程度的相互联系、相互制约的依存关系。例如，市场商品价格与其需求量之间、家庭的工资收入与日常生活之间，都在一定程度上存在相互依存的关系。所有这些依存关系都可以通过统计分组分析出因素与结果之间的变动规律。

【例 3-2】某企业在不同生产规模情况下，产量与产品单位成本的数据如表 3-2 所示。

表 3-2 某企业产量与单位成本的情况

产量/万件	单位成本/(元/件)
0~2	12
2~4	11.5
4~6	10.8
6~8	10
>8	8.9

从表 3-2 中可以看出，产量越大，产品单位成本越小，表明产品单位成本随产量扩大而降低。这种依存关系只有通过分组才可以观察到。

统计分组在三个方面的作用是相互联系、相互补充的，同时也可以看出统计分组在统计研究中的重要地位。统计分组是后续统计工作的基石，应用于统计工作全过程，是统计研究的基本方法之一。

二、选择分组标志

统计分组中的关键问题在于选择分组标志。分组标志就是将统计总体区分为若干组成部分的根据。社会经济现象一般都有许多不同的标志。对同一总体进行分组会有多种选择，为确保分组后的各组能够正确反映事物内部的规律性，选择分组标志时，应遵循以下原则。

▶ 1. 根据统计研究的目的与任务选择分组标志

在对社会经济现象进行研究时，不同的研究任务要选择不同的分组标志进行分组。例如，以全国工业企业为总体进行研究时，这个研究对象就有很多标志，如经济类型、固定资产原值、职工人数、所属行业等。在具体研究过程中到底应该采用哪种标志进行分组，要根据研究的目的来确定。如果研究的目的是要分析不同经济类型的企业在总体中的构成，那么就要选择经济类型作为分组标志；如果要研究工业企业规模构成状况，则可以选择产值、固定资产原值等作为分组标志。

▶ 2. 选择最能反映现象本质特征的标志作为分组标志

由于社会经济现象复杂多样，因此在选择分组标志时有多种选择，这就需要根据被研究对象的特征，选择最主要的、最能反映事物本质特征的标志进行分组。例如，研究某企业职工生活水平高低情况，可以用职工的工资水平作为分组标志，也可以用职工家庭成员人均收入水平作为分组标志。相对来说，职工家庭成员人均收入水平更能反映职工生活水平的高低，更能反映现象的本质特征，因为即使某一职工工资水平较高，但如果他赡养的人口数很多的话，其家庭生活水平也不会很高。在进行统计分组时，就要选择其中最能反映问题本质特征的标志即职工家庭成员人均收入进行分组，这样能够使我们对所研究的对象有一个正确的认识。

▶ 3. 根据现象所处的历史时期来选择标志

社会经济现象是随着时间、地点的变化而变化的。同一个标志在过去某个时期是适用的，现在不一定适用；在这个场合适用，在另一场合不一定适用。因此，即使是研究同类现象，也要视具体时间、地点、条件的不同而选择不同的分组标志。例如，反映企业规模的标志有职工人数、生产能力、固定资产价值等，究竟应选择哪个标志作为分组标志，需视具体条件而定。在技术不发达或劳动密集的条件下，适合用职工人数来表示企业规模；反之，在技术进步或技术装备比较先进的条件下，采用固定资产价值或生产能力来表示企业规模会更恰当、更切合实际。

尤其需要注意的是，在选择分组标志时，还要遵循穷尽性和互斥性两个原则。穷尽性原则是指统计分组必须保证总体的每一个单位都能归入其中的一个组，各个组的单位数之和等于总体单位总数，总体的指标必须是各个单位相应标志的综合；互斥性原则是指统计分组必须保证总体的每一个单位只能属于其中的一个组，不能出现重复统计的现象，否则必然影响统计资料的真实性。

三、统计分组的方法

统计分组要求将总体内标志表现不同的总体单位分开，使标志表现相同或相近的总体单位归属在同一组。因此，分组标志一经选定，就要突出总体在这一标志下的性质差异或数量差异，即在分组标志范围内，划分各相邻组间的性质界限和数量界限。

(一) 按分组标志性质不同分组

按分组标志性质不同，统计分组方法分为品质标志分组和数量标志分组。

▶ 1. 按品质标志分组

按品质标志分组就是选择反映事物属性差异的品质标志为分组标志，并在品质标志的变异范围内划定各组界限，将总体划分成若干个性质不同的组成部分。

【例 3-3】2015 年年末部分地区的人口见表 3-3，我国 2006—2015 年县级市数按年份进行的分组见表 3-4。

表 3-3 2015 年年末部分地区的人口

单位：万人

地 区	人 口 数	地 区	人 口 数
北京	2 171	湖北	5 852
天津	1 547	湖南	6 783
河北	7 425	广东	10 849
山西	3 664	广西	4 796
内蒙古	2 511	海南	911
辽宁	4 382	重庆	3 017
吉林	2 753	四川	8 204
黑龙江	3 812	贵州	3 530
上海	2 415	云南	4 742
江苏	7 976	西藏	324
浙江	5 539	陕西	3 793
安徽	6 144	甘肃	2 600
福建	3 839	青海	588
江西	4 566	宁夏	668
山东	9 847	新疆	2 360

资料来源：中华人民共和国国家统计局 2016 年国家数据。

表 3-4 我国 2006—2015 年县级市数

单位：个

年 份	县 级 市 数	年 份	县 级 市 数
2006	361	2011	370
2007	361	2012	367
2008	368	2013	368
2009	368	2014	368
2010	369	2015	369

资料来源：中华人民共和国国家统计局 2016 年国家数据。

▶ 2. 按数量标志分组

按数量标志进行分组，就是根据统计研究的目的，选择反映事物数量差异的数量标志作为分组标志，在数量标志值的变异范围内划定各组的数量界限，将总体划分为性质不同的若干个部分或组别，结果形成变量数列。根据数量标志值的多少不同，可以划分为以下类型。

(1) 单项式分组和组距式分组。单项式分组是指分组后，每组组别只有一个标志值的统计分组。例如，某城市家庭按人口多少进行分组，可以分为(1, 2, 3, 4, 5)，按每个家庭汽车拥有量(1, 2, 3, 4, 5)分组时，由于其中标志值变动的范围比较小，因而均可采用单项式分组。

组距式分组是指分组后，每组组别对应多个标志值而形成区间的统计分组，其中区间的距离称为组距，故称组距式分组。对于离散变量，如果变量值的变动幅度很大，变量值个数很多，则可进行组距式分组。例如，在百分制考试中，学生按考试成绩分组；我国人口按年龄进行分组等，均适合采用组距式分组。

对连续变量进行分组，由于不能一一列举其变量值，只能采用组距式的分组方式，如企业按产值、营业利润、劳动生产率或工资等标志进行的分组。

(2) 等距分组和不等距分组。等距分组是各组保持相等的组距，即各组标志值的变动都限于相同的范围。不等距分组即各组组距不相等，也称为异距分组。某地区人口按年龄分组编制的等距数列和异距数列分别见表 3-5 和表 3-6。

表 3-5 某地区人口年龄构成(等距数列)

按年龄分组/岁	人口数/万人	比重/%
10 以下	50	2.538
10~20	180	9.137
20~30	320	16.244
30~40	520	26.396
40~50	450	22.843
50~60	300	15.228
60 以上	150	7.614
合计	1 970	100.000

表 3-6 某地区人口年龄构成(异距数列)

按年龄分组/岁	人口数/万人	比重/%
1 以下(婴儿组)	30	1.5
1~3(幼儿组)	100	5.0
3~7(学龄前儿童组)	220	11.0
7~18(青少年组)	380	19.0
18~35(青年组)	620	31.0
35~60(中年组)	490	24.5
60 以上(老年组)	160	8.0
合计	2 000	100.0

在表 3-6 中，人口按年龄进行异距分组，将全部人口划分为婴儿组、幼儿组、学龄前儿童组、青少年组、青年组、中年组和老年组 7 个类别。与表 3-5 相比，表 3-6 更清楚地显示出人口的年龄构成，当然通过这个案例我们也发现，按品质分组和按数量分组有时也不是完全割裂开的。

统计分组时采用等距分组还是不等距分组，取决于研究对象的性质和特点。在标志值变动比较均匀的情况下宜采用等距分组。等距分组便于各组单位数和标志值直接比较，也便于计算各项综合指标。在标志值变动很不均匀时，按一定的比例关系变化，或取值呈现极端偏斜分布状态时，宜采用不等距分组，此时更能说明现象的本质特征。

(二) 按分组标志的多少分组

根据分组标志的多少，统计分组方法分为简单分组和复合分组。

▶ 1. 简单分组

简单分组是指按照一个标志进行分组，表 3-7 所示为某班级按分数对期末考试成绩进行分组。

表 3-7 某班级按分数对期末考试成绩进行分组

成绩/分	人数/人
60 以下	5
60~70	12
70~80	20
80~90	14
90 以上	3
合计	54

▶ 2. 复合分组

复合分组是指对同一个总体，把两个或两个以上标志层叠起来进行分组，表 3-8 所示为按职称、年龄、性别对某高校教师进行复核分组，这里用三个标志的层叠进行分组。

表 3-8 某高校教师的复合分组

职称(第一标志)	年龄(第二标志)	性别(第三标志)
高级职称	45 岁以下	男
		女
	45 岁以上	男
		女
非高级职称	45 岁以下	男
		女
	45 岁以上	男
		女

任务三 分配数列

分配数列是统计整理结果的一种重要表现形式，它可以表明总体单位在各组间的分布特征、结构状况，在此基础上进一步来研究标志的构成、平均水平及其变动规律。由于分组标志的性质不同，分配数列的编制方法当然也是不同的。

一、分配数列的概念

分配数列是在统计分组的基础上，将总体的所有单位按组归类整理，并按一定顺序排列而形成的总体中各个单位在各组间的分布，又称分布数列或次数分配。

在分配数列中，分布在各个组的总体单位数叫作次数或频数。各组次数与总次数之比称为比率或频率。频率分布能够正确反映现象在总体中的分布特征、结构状况，是分析总体特征及其变动规律的重要手段。

由此可见，分配数列有两个组成要素：一个是分组；另一个是次数或频率。表 3-9 所示为某企业职工学历结构情况。

表 3-9 某企业职工学历结构情况

文化程度	职工人数/人	占总人数的比重/%
本科及本科以上	250	12.5
大专及大专以上	350	17.5
中专及高中	800	40.0
初中	600	30.0
合计	2000	100.0

表 3-9 中，第一列为各组指标名称；第二列为次数或频数；第三列为比重或频率。

二、分配数列的种类

根据分组标志的不同，分配数列分为品质分配数列和变量分配数列两种。

(一) 品质分配数列

按品质标志分组形成的分配数列叫作品质分配数列，简称品质数列，也称属性分布数列。表 3-10 所示为 2015 年某地区国内生产总值及构成情况。

表 3-10 2015 年某地区国内生产总值及构成情况

产业	国内生产总值	
	绝对数/亿元	相对数/%
第一产业	22 436.0	11.6
第二产业	93 162.0	48.4
第三产业	76 972.0	40.0
合计	192 570.0	100.0

(二) 变量分配数列

按数量标志分组所编制的分配数列叫作变量分配数列，简称变量数列。

在分组过程中，根据变量值取值范围不同，分组的形式可以分为单项式分组和组距式分组。变量数列按变量的表示方法和分组方法不同，可以分为单项式分配数列和组距式分配数列两种。

▶ 1. 单项式分配数列

单项式分配数列是指将每一变量值列为一组形成的数列，即每一组只包含一个变量值，按单项式分组所编制的变量数列。这种分组形式只适用于离散变量，而且要求在离散变量的变动范围较小、变量值个数较少时使用。表 3-11 所示为某小区家庭按照儿童数分组。

表 3-11 某小区家庭按照儿童数分组

家庭按照儿童数分组/个	家庭数/户
0	20
1	60
2	150
3	90
4	40
合计	360

▶ 2. 组距式分配数列

组距式分配数列是以标志值变动的范围作为一组的分组，即组距式分组所形成的变量数列，适用于所有的连续变量和取值范围较大的离散型变量。表 3-12 所示为某班同学“统计学”成绩分布。

表 3-12 某班同学“统计学”成绩分布

成 绩	人 数/人	比 例/%
60 以下	2	2.50
60~69	15	18.75
70~79	36	45.00
80~89	20	25.00
90 以上	7	8.75
合计	80	100.00

三、变量分配数列的编制

对于品质数列来讲，如果分组标志选择得好，分组标准定得恰当，则事物性质的差异表现得就比较明确，总体中各组也容易划分。在编制品质数列时，只要按规定的分组标准将总体单位按组归类整理即可。品质数列一般来说比较稳定，通常能准确反映总体的分布

特征。在这里，我们只对变量分配数列的编制做重点介绍。

► 1. 单项式数列的编制

在编制单项式数列时，一般将调查所得资料按照数值由小到大顺序排列；然后确定各组的变量值和组数，一般有多少个变量值就有多少组；最后汇总出各变量值出现的次数，编制单项式数列。

由于单项式数列每组只有一个变量值，各组之间的界限划分也非常明确，因此编制出的数列也很稳定。

单项式数列一般只能用在变量值变化幅度不大的离散变量中。当变量值变动幅度较大时，如果采用单项式数列，组数太多，不便于分析问题，也难以反映总体单位在各组的分布趋势，这就需要采用组距式数列。连续型变量只能编制组距式变量数列。

【例 3-4】抽取某班级 20 名同学调查他们的年龄，试编制单项式变量数列。

19 18 20 20 20 19 18 19 19 21
20 20 19 20 20 21 19 20 20 19

具体步骤如下。

(1) 将总体单位标志值从小到大排列起来：

18 18 19 19 19 19 19 19 19 20
20 20 20 20 20 20 20 20 21 21

(2) 以每一岁作为一个组，以每一组出现的次数为各组次数，编制单项式数列，如表 3-13 所示。

表 3-13 某班级 20 名学生年龄统计表

学生年龄/岁	频 数/人	频 率/%
18	2	10
19	7	35
20	9	45
21	2	10
合计	20	100

在此案例中，学生年龄范围是 18~21 岁，数据跨度较小。如果资料数据跨度较大，那么用单项式变量数列进行编制，组数就会比较多，这样就很难看出数据分布的特点，在这种情况下就需要使用组距式变量数列。

► 2. 组距式变量数列

连续型变量和变量值跨度较大的离散型变量适合编制组距式变量数列。组距式变量数列不再是一个变量值代表一个组，而是将整个变量值一次划分为几个区间，一个区间为一组，由这些组所包含的次数所组成的变量数列构成组距数列。

编制这种数列涉及的问题相对较多，主要有全距、组限、组距、组数、组中值、等距数列和异距数列等。

(1) 全距：全距是指总体中最大标志值与最小标志值之差。

(2) 组限：组限是指组距数列中，每个组两端的变量值。其中每组的最大值叫上限，

每组的最小值叫下限，统称为组限。

如果各组的组限都齐全，称为封闭组。如果组限最小组缺下限或最大组缺上限，称为开口组。缺下限的开口组，组距数列的第一组用“×××以下”表示；缺上限的开口组，组距数列的最后一组用“×××以上”来表示。

划分连续型变量的组限时，依据“重叠分组”和“上组限不在组内”原则。相邻两组的上下限应用同一变量值表示，即相邻两组的上下限必须重叠。一般应把重叠的数值归入下限那一组，即“上组限不在组内”原则。

划分离散型变量组限时，相邻两组的上下限可用两个连续自然数表示，也可用“重叠分组”。

(3) 组距：在组距式分组中，组距是各组上下限之间的举例，也就是各组最大标志值和最小标志值之差，即组距=上限-下限。按数量标志进行组距式分组，还可分为等距分组和不等距(或称为异距)分组。

等距分组就是标志值的各组保持相等的组距。等距分组便于绘制统计图，也便于进行各类运算。标志值变动比较均匀的情况下，建议采用等距分组。

异距分组就是各组的组距不相等。在标志值分布很不均匀，如急剧增长、下降或变动幅度很大时，应考虑采用异距分组。

(4) 组数：组数是指根据一定的组距划分的区间的数目。组数与组距呈反比例关系。组数越多，则组距越小；反之，组距越大，则组数越少。对于组数和组距，先确定哪个，应视具体情况而定。

(5) 组中值：组中值是指各组中点位置所对应的变量值。

对于封闭组，组中值的计算公式为

$$\text{组中值} = \frac{\text{本组上限} + \text{本组下限}}{2}$$

对于开口组，组中值的计算公式为

$$\text{缺上限组的组中值} = \text{本组下限} + \frac{\text{邻组组距}}{2}$$

$$\text{缺下限组的组中值} = \text{本组上限} - \frac{\text{邻组组距}}{2}$$

在计算平均指标或进行其他统计分析时，常以组中值来代表各组标志值的平均水平。当各组标志值均匀分布时，组中值代表各组标志值的水平，其代表性就高。

组距式变量数列的编制比较复杂，下面结合实例具体说明。

【例 3-5】根据下述资料编制变量数列。

某单位 30 名工人的月生产零件数(单位为件)如下：

1 060	840	1 100	910	1 090	910
1 110	1 070	990	940	1 190	870
1 180	970	1 030	1 060	850	1 060
1 010	1 050	960	1 050	1 070	1 210
1 050	950	1 060	1 280	1 110	1 010

1. 将原始资料按大小顺序排列

840	850	870	910	910	940
-----	-----	-----	-----	-----	-----

950	960	970	990	1 010	1 010
1 030	1 050	1 050	1 050	1 060	1 060
1 060	1 060	1 070	1 070	1 090	1 100
1 110	1 110	1 180	1 190	1 210	1 280

2. 确定组距和组数

(1) 在排序好的数字中, 找出最大值和最小值, 计算全距(R)=最大值-最小值=1 280-840=440(件)。

(2) 确定组数(k)和组距(i)

$$k=1+3.3221\log n \quad (3-1)$$

≈ 5.91 (由于 k 为组数, 故取整, 即 $k=6$ 组)

$$i=\frac{R}{k} \quad (3-2)$$

$$=\frac{440}{6}\approx 73.33(i \text{ 为组限, 取 } 75, \text{ 非四舍五入})$$

式(3-1)称为斯德吉斯(H. A. Sturges)组数公式。式(3-1)和式(3-2)中, n 为数据的个数, R 为全距。不过实际中也可以视情况先确定组距或组数, 再根据式(3-2)确定另一个。先确定哪一个, 应考虑以下两点: ①要尽量能反映总体单位的分布情况及总体单位的集中趋势; ②要尽可能区分组与组在性质上的差异。

3. 确定组限

确定组限时, 应注意以下问题。

(1) 最小组的下限(起点值)应低于最小变量值, 最大组的上限(终点值)应高于最大变量值, 如图 3-1 所示, 否则会出现有的变量值无法被归入组的情况。



图 3-1 组限确定示意图

(2) 组限的确定应有利于表现总体分布的特点, 应反映出事物质的变化。

(3) 为了方便计算, 组限应尽可能取整, 最好是 5 或 10 的整倍数。

(4) 由于变量有连续型变量和离散型变量两种, 其组限的确定方法是不同。

连续型变量采用“重叠分组”。划分离散型变量组限时, 相邻两组的上下限可用两个连续自然数表示, 也可用“重叠分组”。

4. 编制组距式变量数列

计算表 3-14 中各组的频数和频率, 编制 30 名工人月生产零件数分组表。

表 3-14 某工厂 30 名工人月生产零件数分组表

按生产零件数分组/件	频 数/人	频 率/%
910 以下	3	10.00
910~985	6	20.00

续表

按生产零件数分组/件	频 数/人	频 率/%
985~1 060	7	23.33
1 060~1 135	10	33.33
1 135~1 210	2	6.67
1 210 以上	2	6.67
合计	30	100.00

在变量数列中，我们可以通过频数的合计来检验变量是否不重不漏地被归入各组，其次各组的频率都是介于 0 和 1 之间的，且各组频率之和等于 1。

四、计算累计频数和累计频率数列

为了更详细地认识变量的分布特征，还可以编制累计频数和累计频率数列，可分为向上累计频数(频率)和向下累计频数(频率)两种。

向上累计频数的意义：由变量值小的组向变量值大的组累计各组的频数或频率，称为向上累计频数或向上累计频率。累计数说明小于该组下限的各组的频数(频率)之和。

向下累计频数的意义：由变量值大的组向变量值小的组累计各组的频数或频率，称为向下累计频数或向下累计频率。累计数说明大于及等于该组下限的各组的频数(频率)之和。

依据表 3-14，编制有累计频数频率的分配数列，见表 3-15。

表 3-15 某工厂 30 名工人月生产零件数分组表(含累计频数/频率)

按生产零件数分组/件	频 数/人	频 率/%	向上累计		向下累计	
			频 数/人	频 率/%	频 数/人	频 率/%
910 以下	3	10.00	3	10.00	30	100.00
910~985	6	20.00	9	30.00	27	90.00
985~1 060	7	23.33	16	53.33	21	70.00
1 060~1 135	10	33.33	26	86.66	14	46.67
1 135~1 210	2	6.67	28	93.33	4	13.34
1 210 以上	2	6.67	30	100.00	2	6.67
合计	30	100.00	—	—	—	—

表 3-15 中的数据说明，月生产零件数在 1 060 以下的有 16 人，占总数的 53.33%，即月生产零件数在 1 060 以上的有 14 人，占总数的 46.67%，依此类推。

五、频数分布的主要类型

由于社会经济现象性质各不相同，所以频数分布也不尽相同。频数分布主要有钟形分布、U 形分布和 J 形分布三种类型。

▶ 1. 钟形分布

钟形分布是指靠近两端的变量值分配次数较少，中间变量值分配次数较多，绘制成的曲线图形状宛如一口古钟。

钟形分布是客观现象分布中最常见的分布，其突出特征是“中间大，两头小”。如果钟形分布的中间变量值次数最多，两侧变量值分配的次数随着其与中间变量值距离的增大而渐次减少，并围绕中心变量值两侧呈完全对称分布，则称为对称分布。例如，学生成绩的分布、单位面积的农产品产量等。

许多客观现象总体都趋近于正态分布。中心变量值两侧的变量值次数分布不对称的称为非对称分布或偏态分布，通常有左偏态和右偏态两种。

▶ 2. U形分布

U形分布的特征与钟形分布特征恰恰相反，靠近中间的变量值分布次数少，靠近两端的变量值分布次数多，分布特征是“两头大，中间小”，绘成的曲线图形如英文字母“U”。例如，人口在不同年龄上的死亡率一般近似地表现为U形分布。

▶ 3. J形分布

J形分布有正反两种情况：次数随变量值增大而增多时所绘成的曲线图形如英文字母“J”，称为正J形分布；次数随变量值增大而减少时所绘成的曲线图犹如反写的英文字母“J”，称为反J形分布。例如，商品供给量随着价格的提高而不断增加，使供给曲线呈正J形分布；人口总体按年龄大小的分布一般呈反J形分布。

频数分布图形如图 3-2 所示。

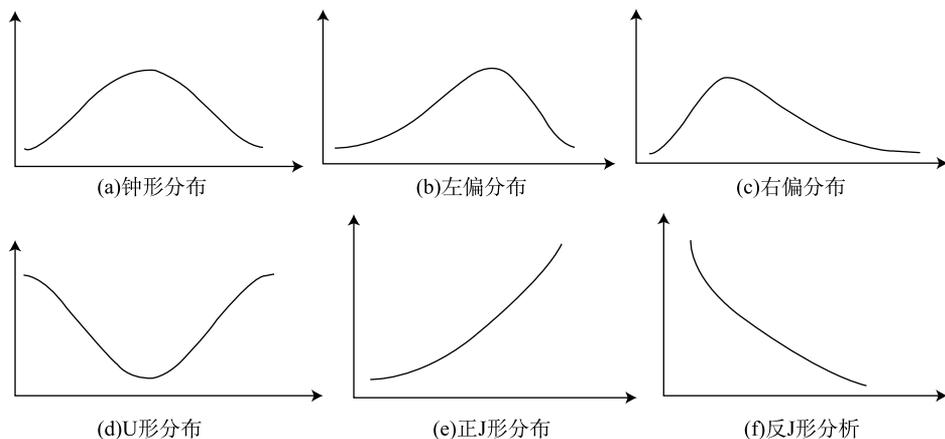


图 3-2 频数分布图形

任务四 统计数据的显示

经过整理的统计资料往往通过统计表和统计图显示出来，因此，统计表和统计图成为显示统计数据的重要工具。

一、统计表

► 1. 统计表的构成

统计表就是由纵横交叉的直线所组成的表格来显示统计数据的表格。统计表的应用范围极其广泛，是表现统计资料最常用的形式。其主要优点是：使统计资料条理化，更清晰地表述统计数据之间的相互联系；使统计数据的显示简明易懂；便于计算和比较表内的各项统计指标，并易于检查数字的完整性和正确性。

统计表的构成可以从表式和内容两个方面来认识。从统计表的形式来看，统计表由总标题、横行标题、纵栏标题和统计数字 4 个要素构成，如图 3-3 所示。

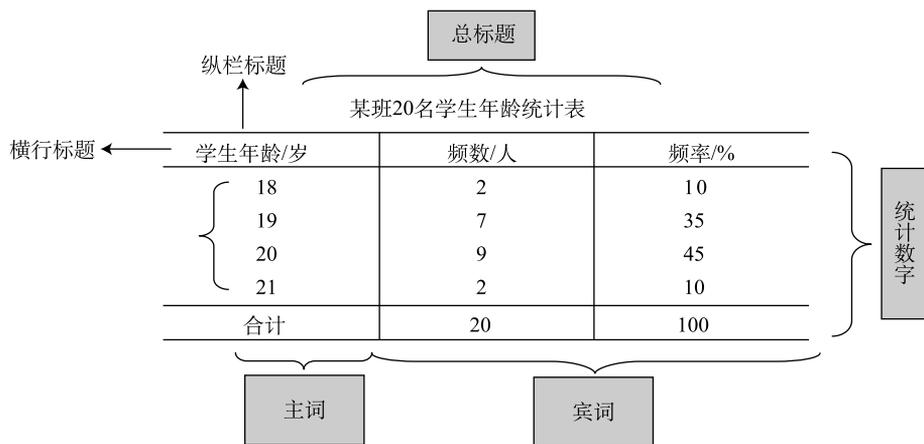


图 3-3 统计表的形式

总标题是统计表的名称，用以概括说明整个表的内容，多数情况要包括总体的时间和空间限制，一般位于表的上方中央。

横行标题是横行内容的名称，通常用来说明总体及其各组的名称，通常也称为主词，一般列在表的左边。

纵栏标题是纵栏内容的名称，通常用来表示反映总体及其各组成部分数量特征的统计指标的名称，通常也称为宾词，一般列在表内的上方。

统计数字是各项指标的具体数值，内容由横行标题和纵栏标题所限定，其数字可以是绝对数、相对数或平均数。

另外，为了补充统计表中未说明的问题，统计表往往还附有一些说明，包括资料来源、指标计算方法、填报单位、填表人、填表日期等。

从统计表的内容来看，由两部分组成，一部分是主词；另一部分是宾词。主词是统计表的主体，也是统计表所要说明的对象。主词通常用横行标题来表示。宾词是说明主词的各项指标，一般由纵栏标题和统计数字所组成。

统计表的主词和宾词的位置一般如图 3-3 所示，但不是固定不变的，有时为了阅读方便，可以将主词和宾词的位置互换。

► 2. 统计表的分类

统计表按对总体分组的情况不同，可以分为简单表、分组表和复合表。

简单表是指对统计总体未做任何分组，仅按单位名称或时间顺序排列而成的统计表，如表 3-16 所示。

表 3-16 中国主要河流基本情况

名 称	流域面积/平方千米	河长/千米	年径流量/ 10^9 米 ³
长江	1 808 500	6 300	9 513
黄河	752 443	5 464	661
松花江	557 180	2 308	762
辽河	228 960	1 390	148
珠江	453 690	2 214	3 338
海河	263 631	1 090	228
淮河	269 283	1 000	622

分组表又称简单分组表，是指对统计总体仅按一个标志进行分组而形成的统计表，如表 3-17 所示。利用分组表可以深入分析现象的内部结构和现象间的相互依存关系。

表 3-17 某企业职工学历构成情况

文 化 程 度	职工人数/人	占总人数的比重/%
大专及大专以上	350	17.5
中专及高中	800	40.0
初中	600	30.0
小学及小学以下	250	12.5
合计	2 000	100.0

复合表又称复合分组表，是指对统计总体按两个或两个以上标志进行层叠分组而形成的统计表，如表 3-18 所示。

表 3-18 2016 年某地区高等学校普通本、专科部分学科招生人数 单位：人

项 目	本 科	专 科	合 计
经济学	152 592	116 181	268 773
法学	110 019	86 176	196 195
文学	470 022	346 900	816 922
工学	798 106	1 194 320	1 992 426
农学	47 312	52 708	100 020
医学	155 242	224 841	380 083
管理学	411 464	662 161	1 073 625
合计	2 144 757	2 683 287	4 828 044

► 3. 编制统计表的要求

为了使统计表能够科学地反映被研究对象的数量特征，同时为了使统计表美观和标准

化，编制统计表时要遵循科学、实用、简练、美观的原则，还要符合以下要求。

(1) 统计表的标题、项目、指标要简明扼要，应写明资料所属的时间和空间范围，使人一目了然，便于分析。如果指标的计量单位只有一个，则通常列在表的右上角，如果计量单位较多，则列在相应的指标栏内。

(2) 统计表的纵栏、横行的排列要尽量反映出内容方面的逻辑关系，如表 3-19 所示。

表 3-19 某地区 2016 年国民生产总值和职工人数统计表

项 目	国民生产总值/万元	职工人数/人
国有经济	大型	13 030
	中型	9 140
	小型	5 021
集体经济	大型	8 067
	中型	6 160
	小型	4 600
外商投资经济	大型	7 500
	中型	6 210
	小型	5 400
其他经济	大型	5 800
	中型	4 510
	小型	4 016

(3) 当统计表的栏目较多时，可进行编号，以说明其相互关系。主词栏和计量单位栏常用甲、乙、丙等文字编号，宾词栏常用 1、2、3 等数字编号。

(4) 表中的合计栏可以排在前面，也可以排在最后，如果只列出其中部分项目，则合计栏必须排在前面。

(5) 表中的统计数字应书写工整、字迹清晰；数字应填写整齐，数位对准。数字为零时要写出“0”来，不应填写数字的空格用“—”线表示；未发生的数字空着不填；估算的数字应在表下说明；无法取得的资料用“……”号表示；如果某项数字与邻项数字相同，则仍应填写数字，不得用“同上”“同左”等字样或符号代替。

(6) 国际上规范的统计表是“三线表”，表的上下两端用粗线，左右两边不封口，纵栏之间用细线分开，横行之间可以不加线。

(7) 统计表的资料来源及其他需要说明的问题可在表下加以注明。例如，统计资料的来源、填表时间、制表人、审核人等。

(8) 统计表中的数字资料都要注明计量单位。计量单位应按统计制度的规定填写，不得另设不同的计量单位。

为使统计表阅读方便，计算单位应按如下方法表示：当各指标数都以同一单位计量时，就将计量单位写在统计表的右上角；当同栏指标数值以同一单位计量时，而各栏的计量单位不同时，则应将单位写在各纵栏标题的下方或右方；当同行统计资料以同一单位计量，而各行的计量单位不同时，则可在横行标题后添加计量单位栏，用以标明各行的计量单位。

二、统计图

统计图是人们用来展示统计资料的另一种常用形式，是根据整理过的统计资料，运用几何图形或具体的形象来表现被研究对象数量关系和数量特征的图形。与统计表相比，统计图更为鲜明、醒目、生动、直观。常用的统计图有柱形图、直方图、折线图、圆形图、散点图等。

► 1. 柱形图

柱状图分为单式柱形图和复式柱形图。

(1) 单式柱形图。根据表 3-20 绘制单式柱形图，如图 3-4 所示。

表 3-20 某班 20 名学生年龄统计表

学生年龄/岁	频 数/人	频 率/%
18	2	10
19	7	35
20	9	45
21	2	10
合计	20	100

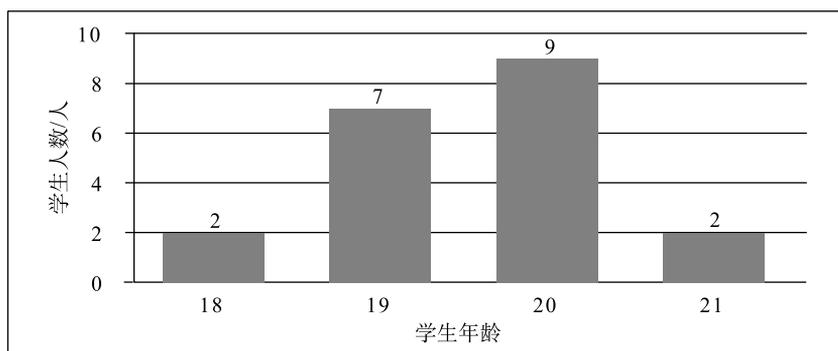


图 3-4 某班 20 名学生年龄柱形图

(2) 复式柱形图。根据表 3-21 中的数据资料，为反映某企业各车间的计划产量与实际产量的对比情况，绘制复式柱形图，如图 3-5 所示。

表 3-21 某企业 2016 年产量计划完成情况表

车 间	计划产量/件	实际产量/件
A 车间	166	168
B 车间	194	240
C 车间	240	312
D 车间	250	260
合计	850	980

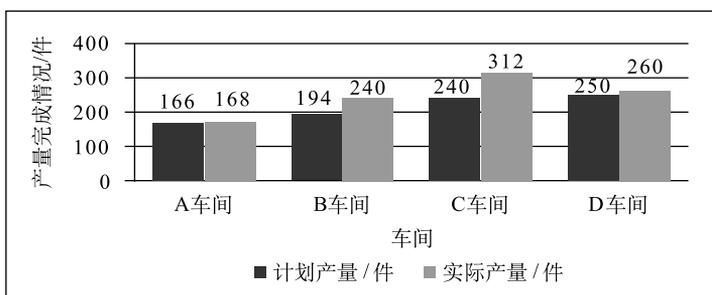


图 3-5 某企业各车间产量完成情况对比图

► 2. 直方图

直方图是用矩形的宽度和高度来表示频数分布的图形。在平面直角坐标系中，横轴表示数据分组，纵轴表示频数或频率，这样各组与相应的频数就形成了一个矩形，即直方图，一般用来表现连续型变量的分布特征。例如，根据表 3-22 绘制频数直方图，如图 3-6 所示。

表 3-22 某工厂 30 名工人月生产零件数分组表

按生产零件数分组/件	频 数/人	频 率/%
910 以下	3	10.00
910~985	6	20.00
985~1 060	7	23.33
1 060~1 135	10	33.33
1 135~1 210	2	6.67
1 210 以上	2	6.67
合计	30	100.00

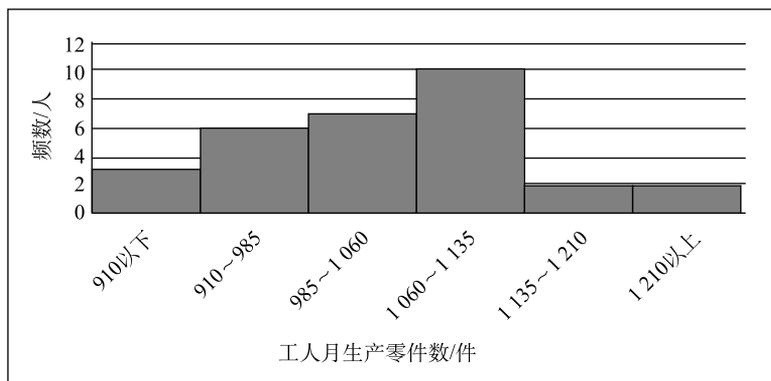


图 3-6 某工厂 30 名工人月生产零件频数直方图

► 3. 折线图

折线图也称频数多边形，是在直方图的基础上把相邻直方形的顶边中点连接成一条折

线，再把折线两端与横轴上直方形两侧延伸的假象组中点相连，就形成了频数分布折线图，也可以用组中值与次数求坐标点连接而成。例如，根据表 3-20 对该班学生年龄绘制折线图，如图 3-7 所示。

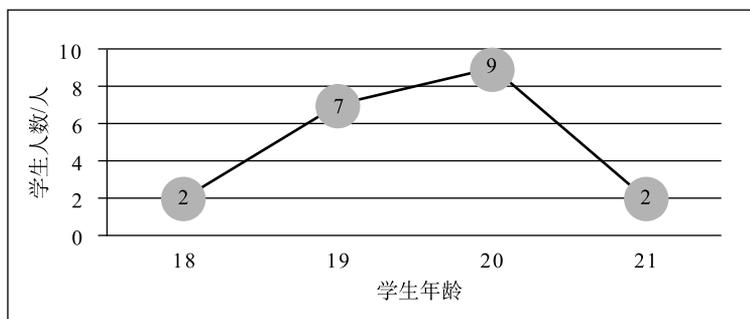


图 3-7 某班 20 名学生年龄结构折线图

► 4. 圆形图

圆形图又称饼图，是以圆的面积或圆内各扇形的面积来表示数值大小或总体内部结构的一种图形。根据表 3-20 绘制圆形图以反映学生年龄结构，如图 3-8 所示。

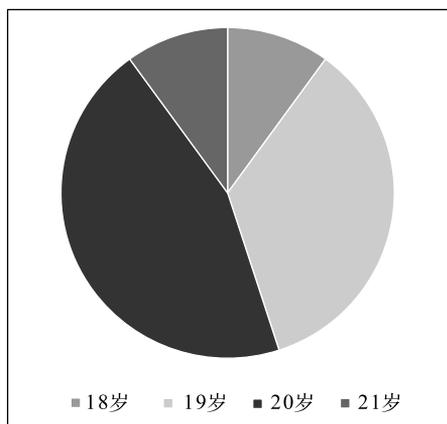


图 3-8 某班 20 名学生年龄结构圆形图

► 5. 散点图

散点图主要用于显示时间序列数据，以反映事物发展变化的规律和趋势。表 3-23 反映了 2011—2016 年我国某省城乡居民家庭的人均收入情况，根据该表绘制散点图，如图 3-9 所示。

从图 3-9 中可以清楚地看出，城乡居民的家庭人均收入逐年提高，而且城镇居民的家庭人均收入高于农村，2011 年以后城乡居民家庭人均收入的差距有扩大的趋势。

表 3-23 2011—2016 年我国某省城乡居民家庭的人均收入情况

年份	2011	2012	2013	2014	2015	2016
城镇居民家庭人均收入/元	2 477.4	3 446.4	4 288.6	4 837.6	5 166.3	5 425.1
农村居民家庭人均收入/元	982.6	1 334	1 671.1	1 926.3	2 092.4	2 462.2

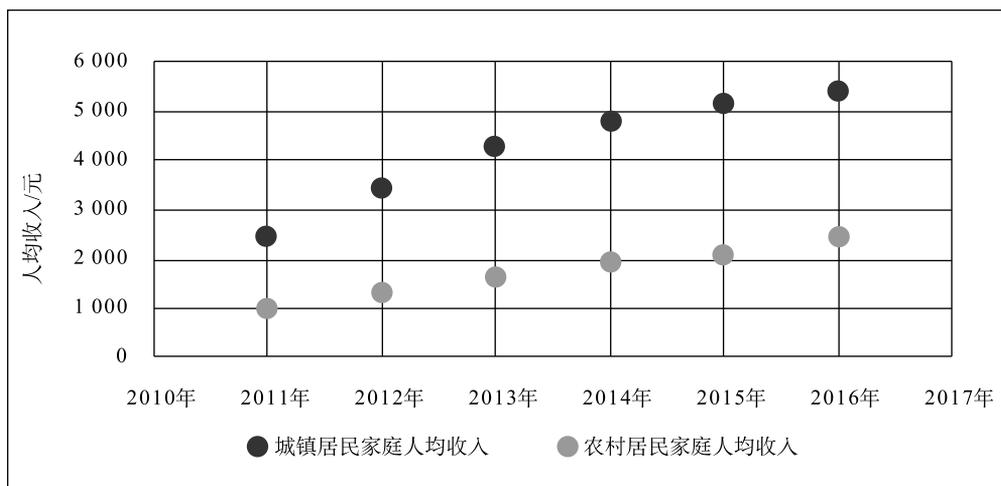


图 3-9 2011—2016 年我国城乡家庭的人均收入散点图

任务五 Excel 在统计整理中的应用

一、利用 FREQUENCY 函数进行统计数据的分组整理

用各种方法取得的统计数据必须经过加工整理，使之系统化、条理化，才能符合统计分析的要求。在 Excel 的统计函数中有一个专门于统计分组的 FREQUENCY 函数，可用于分组、计算频数和频率等。下面举例说明该函数的使用方法。（此处使用的 Excel 为 2010 版）

【例 3-6】某单位 30 名工人的月生产零件数(单位为件)如下：

1 060	840	1 100	910	1 090	910
1 110	1 070	990	940	1 190	870
1 180	970	1 030	1 060	850	1 060
1 010	1 050	960	1 050	1 070	1 210
1 050	950	1 060	1 280	1 110	1 010

现准备将这 30 名工人月生产零件数分为 6 组，分别为 910 以下、910~985、985~1 060、1 060~1 135、1 135~1 210、1 210 以上。

第一步，将 30 名工人的月生产零件数输入 A1~A30 单元格，并选定 C3:C8 单元格作为放置分组结果的区域，如图 3-10 所示。

第二步，从“公式”菜单中选择“fx 插入函数”项，在弹出的对话框的“选择类别”列表中选择“统计”，在“选择函数”列表中选择“FREQUENCY”，按回车键进入 FREQUENCY 函数参数对话框，如图 3-11 所示。

	A	B	C	D
1	1060			
2	840			
3	1100			
4	910			
5	1090			
6	910			
7	1110			
8	1070			
9	990			
10	940			
11	1190			

图 3-10 输入月生产零件数并选定放置分组结果的区域



图 3-11 FREQUENCY 函数参数对话框

第三步，在 FREQUENCY 函数参数对话框中 Data_array 文本框中输入待分组计算频数分布原数据，本例可输入 A1:A30；在 Bins_array 文本框中输入分组标志。FREQUENCY 函数要求按组距的上限分组，不接受非数值字符的分组标志（如“××以下”或“不足××”之类），因此，断开的分组标志可以直接输入各组上限数值，而重叠的分组标志则以各组上限减 1 的方式确定分组标志，这样上限数值自动计入下一组。本例的月产量分段区间为 910 以下、910~985、985~1 060、1 060~1 135、1 135~1 210、1 210 以上，因此可输入 909、984、1 059、1 134、1 209、1 284。由于分组结果要给出一组频数，故必须以数级公式的形式输入，即在输入数据的两端加大括号 {}，各数据之间用分号隔开，即输入 {909; 984; 1 059; 1 134; 1 209; 1 284}。需要注意的是，如果分组变量为连续变量，而且变量值中有小数的话，那么分组标志则应以各组上限减 0.1、减 0.01 或减 0.001 等的方式确定，至于减多少要根据变量值的小数位数确定。

输入完毕，即在对话框中看到频数分布“3；6；7；10；2；2”（后面的 0 表示没有其他）。

第四步，将频数分布“3；6；7；10；2；2”记入指定的 C3:C8 单元格内。

第五步，取得频数分布后，可按图 3-12 所示将横行标题和纵栏标题填写齐全。

	A	B	C	D	E	F	G	H	I
1	1060				向上累计		向下累计		
2	840	按生产零件数分组/件	频数/人	频率/%	频数/人	频率/%	频数/人	频率/%	
3	1100	910以下	3	10	3	10	30	100	
4	910	910~985	6	20	9	30	27	90	
5	1090	985~1060	7	23.33	16	53.33	21	70	
6	910	1060~1135	10	33.33	26	86.66	14	46.67	
7	1110	1135~1210	2	6.67	28	93.33	4	13.34	
8	1070	1210以上	2	6.67	30	100	2	6.67	
9	990	合计	30	100	-	-	-	-	
10	940								

图 3-12 某工厂 30 名工人月生产零件数分布表输入图

第六步，取得频数分布后，再列表计算频率及累计频数和频率。

(1) 人数合计，可单击 C9 单元格，输入“=SUM(C3:C8)”，按回车键得出结果为 30 人(SUM 是求和函数)。

(2) D 列频率，可先单击 D3 单元格，输入“=C3/30*100”(* 是乘法符号，除数要直接输入数字“30”，否则无法使用填充柄功能)，按回车键得出结果为 10%；然后利用填充柄功能按住鼠标左键向下拖动至 D9 单元格放开鼠标，即得出 D4、D9 单元格的频率。

(3) E 列向上累计次数可先单击 E3 单元格，输入“=C3”，再单击 E4 单元格，输入“=E3+C4”，然后利用填充柄功能按住鼠标左键向下拖曳，至 E8 单元格放开鼠标，即得出 E5—E8 单元格的累计次数。F 列引用 E 列公式即可得到累计频率。G 列、H 列可仿照此法计算。

二、利用图表向导绘制统计图

利用例 3-6 整理出的频数分布表，选中 B3:C8 单元格，单击“插入”→“图表”，在“插入图表”对话框的“推荐的图表”选项卡中可以看到系统推荐的统计图(见图 3-13)，当然也可以选择“所有图表”(见图 3-14)，如柱形图、饼形图等，在子图表类型中选择适当的类型，此处以柱形图为例。



图 3-13 “推荐的图表”选项卡



图 3-14 “所有图表”选项卡

单击“确定”按钮即可完成统计图的绘制，如图 3-15 所示。

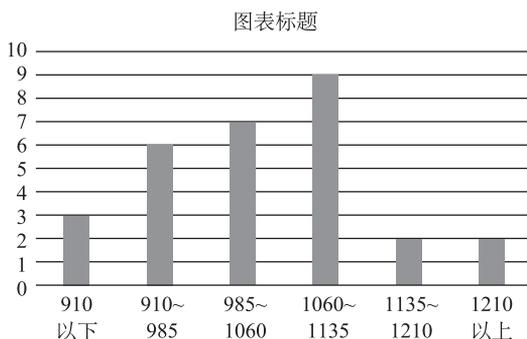


图 3-15 某工厂 30 名工人月生产零件数频数分布

图 3-15 的格式可以调整，如果要消除各个分类间距(各个直方形中间的间距)，则双击直方形打开设置数据系列格式选项，单击“系列选项”，将分类间距调整为 0，如图 3-16 所示。

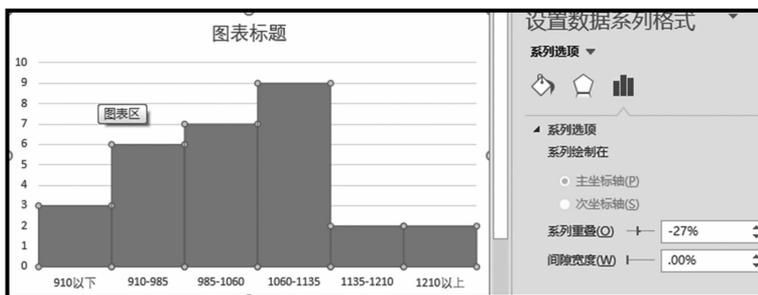


图 3-16 设置数据系列格式分类间距

如果希望将各组数据用不同颜色表示，选择“系列选项”中的“填充与线条”，在“颜色”选项中选择自己喜欢的颜色即可，如图 3-17 所示。

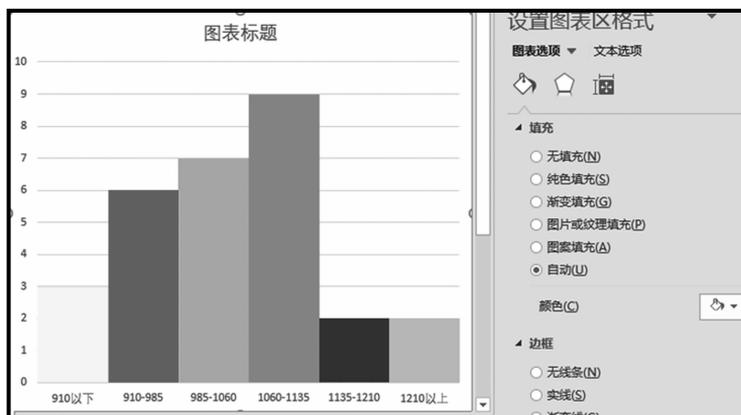


图 3-17 设置数据系列格式颜色填充

最后再加上图表标题、坐标轴标题、图例、趋势线，调整图表大小即可，如图 3-18 所示。

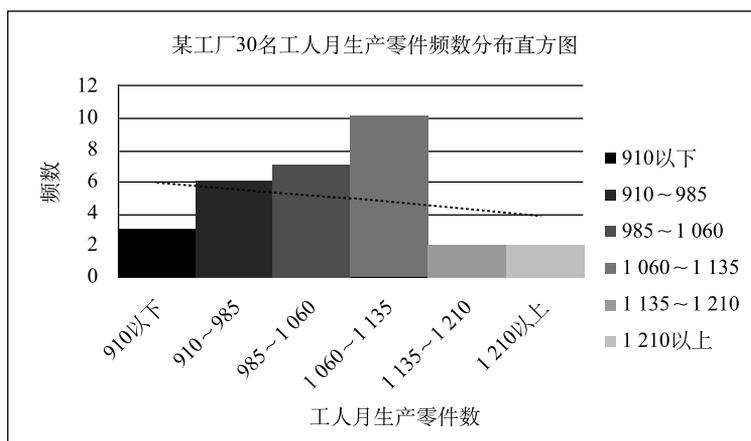


图 3-18 某工厂 30 名工人月生产零件频数分布直方图

利用相似的方法可以制作各种不同的图形，如圆形图，如图 3-19 所示。

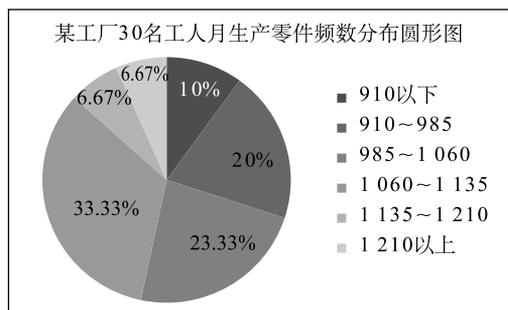


图 3-19 某工厂 30 名工人月生产零件频数分布圆形图

三、利用数据分析工具分组并绘制直方图

Excel 提供了一组数据分析工具——分析工具库，利用该组工具可以简化建立复杂统计或进行工程分析时的步骤。其中有些工具可以用于分组，在产生输出表格的同时，还可以绘制图表。由于在默认的情况下，Excel 并没有安装分析工具库，因此在使用数据分析工具之前，必须先安装分析工具库。方法为单击“文件”→“选项”→“加载项”→“转到”，勾选全部可用加载宏，单击“确定”按钮，之后就可以在“工具菜单”右上端看到“数据分析”选项。

【例 3-7】承例 3-6，准备将这 30 名工人月生产零件数分为 6 组，分别为 910 以下、910~985、985~1 060、1 060~1 135、1 135~1 210、1 210 以上。

第一步，将 30 名工人月生产零件数输入 A2：A31 单元格，如图 3-20 所示。

	A	B	C
1	工人月生产零件数/件		
2	1060		
3	840		
4	1100		
5	910		
6	1090		
7	910		
8	1110		
9	1070		
10	990		
11	940		
12	1190		
13	870		
14	1180		

图 3-20 某工厂 30 名工人月生产零件数据输入界面

第二步，为将样本单位按组归类，还需输入分组标志，但只能按组的“边界值”(组距分组的上限)分组，不能有非数值的字符(如“××以下”“不足××”之类)。本例分为 909、984、1 059、1 134、1 209、1 284 六组，输入 B2：B7 单元格，如图 3-21 所示。

	A	B	C
1	工人月生产零件数/件	工人月生产零件数/件	
2	1060	909	
3	840	984	
4	1100	1059	
5	910	1134	
6	1090	1209	
7	910	1284	
8	1110		
9	1070		
10	990		
11	940		
12	1190		
13	870		
14	1180		

图 3-21 某工厂 30 名工人月生产零件标志输入界面

第三步，在“工具”菜单中单击“数据分析”选项，从对话框的“分析工具”列表中选择“直方图”，如图 3-22 所示，打开“直方图”对话框。

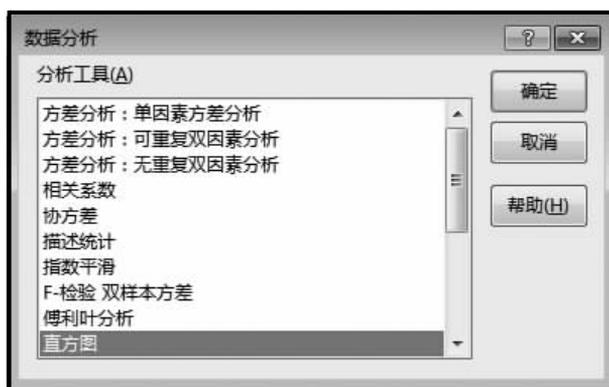


图 3-22 “数据分析”对话框

第四步，在“直方图”对话框的“输入区域”文本框中输入“\$A\$1:\$A\$41”（选中 A1:A31 区域），如图 3-23 所示。由于第 1 行是标志项，还需单击选定“标志”复选框。“接收区域”实际是要求输入分组标志所在的单元格区域，本例可输入“\$B\$1:\$B\$7”（选中 B1:B7 区域）。如果在此框中不输入分组标志所在的区域，系统将在最小值和最大值之间建立一个平滑分布的分组。在“输出区域”文本框中键入输出表左上角的单元格行列号，本例为 C1。如要同时给出次数分布直方图，可单击“图表输出”复选框。如要同时给出“累积%”（通常称“累计频率”），可单击“累积百分率”复选框，系统将在直方图上添加累积频率折线。



图 3-23 “直方图”对话框

第五步，按回车键确认，即在 B 列右侧给出一个 3 列的分组表和一个直方图，如图 3-24 所示。在给定的表和图中，“频率”实际是频数，“累积%”实际是累计频率。



图 3-24 某工厂 30 名工人月生产零件数分组结果及直方图(初图)

在分组表中将多余的“其他”一组删除，修改各组组限，按上述方法将分类间距调节为 0，选择合适的字体并对图形进行适当处理，即可得到合适结果，如图 3-25 所示。

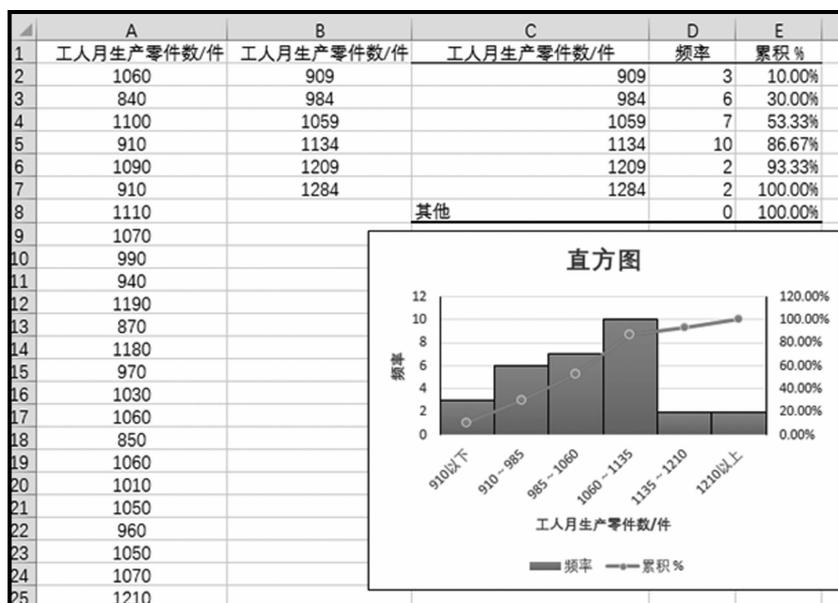


图 3-25 某工厂 30 名工人月生产零件数分组结果及直方图(修改图)

项目小结

统计整理是对统计调查所得到的原始资料进行科学的分类、汇总，或对已经初步加工的资料进行再加工，使之成为系统化、条理化的综合资料，以反映现象总体数量特征的工作过程。统计整理包括对资料的审核、分组、汇总和编制统计表等几个主要环节。

统计分组就是根据统计研究的需要,按照某一标志,将统计总体划分为若干个组成部分。统计分组可以按品质标志分组和按数量标志分组。按数量标志分组又有单项式和组距式两种分组方式。组距式分组要确定好组限、组距和组数等。

频数分布是在统计分组的基础上形成的反映总体单位在各组中分布状况的统计数列,分为品质数列和变量数列两种。变量数列可以编制成频数分布表和分布图。频数分布主要有钟形分布、U形分布和J形分布三种类型。

统计表是把统计数据按照一定的结构顺序,用表格显示的一种形式。统计图是借助几何图形或具体形象显示统计数据的一种形式。最常用的简单统计图形的绘制软件为 Excel,可完成常用的直方图、柱形图、折线图和圆形图等统计图的绘制。

思考与练习

一、单项选择题

- 统计整理主要是针对()进行加工的过程。
A. 综合统计数据 B. 历史数据资料 C. 统计分析数据 D. 原始调查数据
- 统计资料整理的首要环节是()。
A. 编制统计报表 B. 审核汇总资料 C. 审核原始资料 D. 设计整理方案
- ()属于按品质标志分组。
A. 雇员按受教育年限分组 B. 职工按就业领域分组
C. 企业按资产存量分组 D. 住户按人口多寡分组
- 组中值是()。
A. 一个组的上限与下限之差 B. 一个组的上限与下限之间的中点值
C. 一个组的最小值 D. 一个组的最大值
- 将全部变量值依次划分为若干个区间,并将这一区间的变量值作为一组,这样的分组方法称为()。
A. 单变量值分组 B. 组距分组 C. 等距分组 D. 连续分组
- 将某企业职工月收入依次分为 2 000 元以下、2 000~3 000 元、3 000~4 000 元、4 000~5 000 元、5 000 元以上,最后一组的组中值近似为()。
A. 5 000 B. 7 500 C. 5 500 D. 6 500
- 在编制组距数列时,当全距不变的情况下,组距与组数的关系是()。
A. 正比例关系 B. 反比例关系 C. 乘积关系 D. 毫无关系
- 对某地区的全部商业企业按实现的利润额多少进行分组,这种分组属于()。
A. 数量标志分组 B. 属性分组 C. 分组体系 D. 复合分组

二、多项选择题

- 统计分组的作用在于()。
A. 区分现象的类型 B. 反映现象总体的内部结构变化
C. 比较现象间的一般水平 D. 分析现象的变化关系
E. 研究现象之间数量的依存关系

10.0	12.4	17.5	14.5	20.0
6.4	26.0	17.3	14.7	18.2

试根据以上数据进行如下操作：

- (1) 把数据输入 Excel 中。
- (2) 分组并绘制频数分布表。
- (3) 绘制柱状图，并标出每个柱形的数据。
- (4) 分别设置纵、横坐标的标题。
- (5) 更改柱的颜色为红色。
- (6) 更改右边图例系列 1 的名称为 A。
- (7) 把柱状图的各个部分设置为合适的大小，使图形尽量美观。
- (8) 绘制圆形图和折线图，比较哪个图最能直观地反映此数据的规律。

拓展材料

2017 年“双 11”发生了什么？

2017 年 11 月 11 日当天，阿里巴巴交易额冲破 1 亿用了 11 秒钟！3 分 01 秒，突破 100 亿！2016 年，天猫“双 11”成交额突破 100 亿元用时 6 分 58 秒。2017 年，又破了新纪录！2017 年 11 月 11 日上午 9 点 00 分 04 秒，天猫交易额突破 1 000 亿元，打破 2017 年全国社会消费品日均零售额。这一速度也远远超过了 2016 年。2016 年 11 月 11 日 18 时 55 分 36 秒，天猫“双 11”全球狂欢节交易额突破 1 000 亿元，用时不到 19 个小时，而 2017 年只用了 9 个小时。

2017 年 11 月 11 日，阿里巴巴最后的成交额锁定在 1 682 亿元，换句话说，不论是 1—10 月，还是全年，阿里巴巴一天的流水可以轻松进入房地产企业的前十。1 682 亿元还超过 99% 的 A 股上市公司 2016 年全年营业收入，是阿里巴巴 2017 年第二季度营业收入 551 亿元的 3 倍。

“双 11”即将成为覆盖全球 200 多个国家和地区的狂欢日。来自天猫的数据显示，截至 2017 年 11 月 11 日凌晨 1 时，来自日本、美国、澳大利亚、德国和韩国的商品最受消费者喜爱，而来自美国、澳大利亚和新加坡的“剁手党”也在为“双 11”贡献力量。

阿里巴巴为“双 11”晚会请来机器人乐队助兴，这个来自德国的乐队叫作 Compressor-head，由 3 个身高约 1.5 米的机器人组成。在“双 11”晚会现场，它们配合张杰完成了一首歌，并成功抢镜。这个乐队的成员有四条手臂的鼓手 Stickboy“棍子男孩”、78 根指头的吉他手 Fingers“手指头”、有史以来最精确的贝斯手 Bones“骨头”。它们全部由废弃金属打造，由电动气体力学装置控制，并具有 MIDI 信号控制功能。当然，阿里巴巴也想借此告诉大家，2017 年的“双 11”，在挑选货品、导购推荐、客服、设计海报、巡逻机房、管理仓库等领域都有智能机器人入驻，有的工作甚至比人类完成得更好。他们的客服机器人一天能服务 350 万个客户。每个客服机器人“店小蜜”单日对话量都在 1 000 万次以上，全天 24 小时不间断，还能根据用户的购买行为预测问题，快速形成答案。2017 年“双 11”前，“鲁班”AI 设计师智能设计平台已学习数百万设计图像，每秒可生产 8 000 张广告图；机房巡逻员“天巡”能 24 小时巡查数据中心，排查异常，接替了运维人员以往 30% 的重复性工作。目前，阿里巴巴的一个机房已经实现无人作业，由“天巡”看管。有了这些无所不在的

机器人帮手，程序员们表示，比起几年前的帐篷、红牛，现在敷敷面膜就可应对“双 11”，真是改变了不少。2017 年“双 11”，阿里巴巴的某个工作人员和老婆打了个赌，如果有 BUG，就帮她清空购物车，不过，能顶住“小败家们”一波又一波的冲击且没有大崩溃，真心要给阿里巴巴的大数据和云计算技术点个赞。

通过 2017 年的“双 11”，我们可以看到飞速发展的数据处理能力；通过一组组数据、一张张数据表、一幅幅数据图，我们能清晰地感受到现代社会的进步和发展，也能深刻地意识到大数据给人们日常生活带来的翻天覆地的改变。