

项目三

统计整理

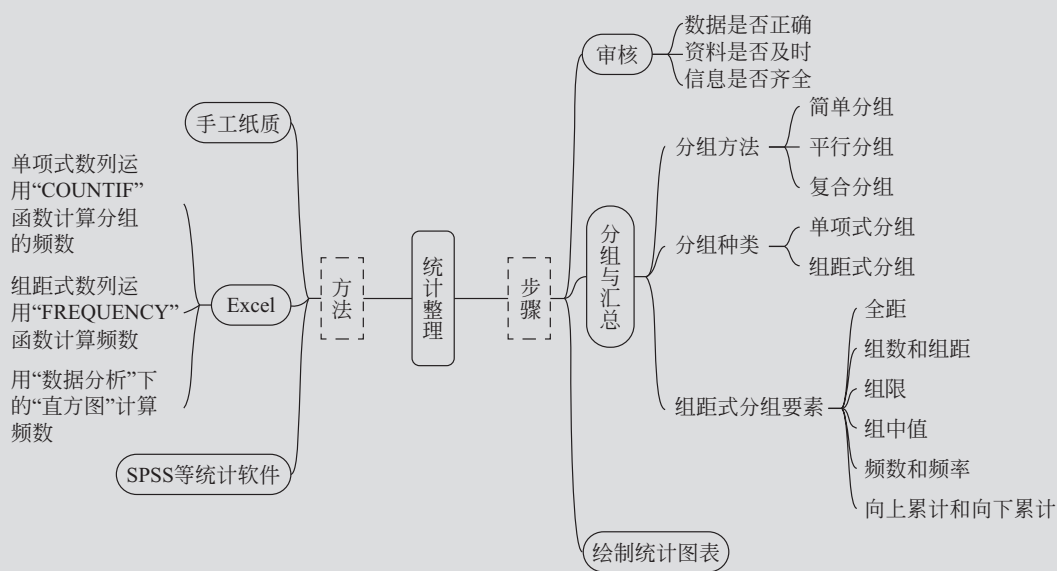
【学习目标】

(1) 知识目标：描述统计整理步骤；描述统计分组的方法；描述分配数列频数、频率等含义，列举分配数列的种类，说明单项式和组距式分配数列差异。

(2) 能力目标：会合理高效地整理统计资料；会编制分配数列；会运用 Excel 计算频数、频率、绘制统计表、统计图。

(3) 素养目标：以统计工作要求养成守时、守信、守岗、守纪、守法“五守”行为习惯；以统计资料质量培养求真实、求精准、求卓越、求创新、求合作“五求”职业精神。

【知识图谱】



【统世经纬】

统计整理相关要求

——摘自《全国统计专业技术高级资格考试大纲(2021 年)》

(1) 掌握数据汇总与整理工作的全过程,按照调查目的和统计调查制度的要求,制订切实可行的数据汇总与整理方案。

(2) 熟悉逻辑审核、相关分析等主要数据质量审核评估方法,能够及时发现并处理基础数据存在的问题。

任务一 初步整理统计资料

【任务描述】

2026年,钱江集团人力资源部组织员工满意度调查,采用问卷星发放问卷,答题结束,共收到问卷2000份,覆盖总部及五大区域分公司。人力资源部需要组织相关工作人员完成2000份问卷审查,以确保问卷的有效性,对有效问卷按员工工作年限进行分组,并利用Excel软件整理答题结果。

【任务准备】

一、统计整理的概念

统计调查得到的数据,只是个别的、分散的、不系统的原始数据,所反映的问题常常是事物的表面现象,不能深刻揭示事物的本质,更不能从量的方面反映事物发展变化的规律性,这就有必要对统计调查所获得的原始资料进行科学整理。只有根据统计研究的目的,运用科学的统计整理方法,对数据进行加工整理,才能发现经济社会现象的数量规律性。

统计整理是统计工作的第二个阶段。它是根据统计研究的任务,对统计调查阶段所搜集到的大量原始资料或次级资料进行加工汇总,使其系统化、条理化、科学化,以反映事物总体综合特征的资料加工过程。

统计整理就是人们对社会经济等现象从感性认识上升到理性认识的过渡阶段,是统计工作中一个十分重要的中间环节,起着承前启后的作用,既是统计调查阶段的继续和深入,又是统计分析阶段的基础。

二、统计整理的步骤

统计资料整理要经过四个步骤:①对资料进行审核;②对资料进行分组与汇总;③将分组或汇总的资料以统计表或统计图的形式呈现;④对统计资料进行系统积累。

1. 审核

审核主要从准确性、及时性、完整性三个方面进行。

(1) 准确性审核主要检查所获取的资料是否真实、可靠,检查的方法有两种:①逻辑检查,从常理角度审核数据是否符合逻辑,有无自相矛盾或不符合实际的地方,内容是否合理,例如,某问题要求女性回答,但是回答的是男性,那么该份问卷不符合逻辑;②计算检查,复核调查表或调查问卷的各项数据有无差错,检查各项指标的计算方法是否正确,计量单位与法定单位是否一致,有关指标间的平衡关系是否得到保持。例如,分组后各组的比重合计应该等于100%,如果不等于,说明要么资料有错,要么计算有错。

(2) 及时性审核主要审查资料是否符合规定的时间,资料报送是否能在调查期限内



微课:统计整理概述



报出。

(3) 完整性审核主要调查项目是否齐全,调查单位是否重复或遗漏。

在发现统计资料数据有差错时,要分不同情况进行处理:属于填报错误的,要通知填表人或填报单位重新填报;属于汇总错误的,应根据情况予以重新核算或修正,并对相关流程进行核查以避免同类错误再次发生。

2. 统计资料分组与汇总

(1) 对统计分组与汇总的理解。统计分组是根据统计研究的目的和客观现象的内在特点,按一定的标志把总体划分成若干不同性质的组,并计算出每一组的频数、频率,根据分组的结果计算总体统计指标。统计分组是基本的统计方法之一,在资料整理和统计分析中都要广泛应用分组。分组的好坏直接关系到统计能否整理出正确的、有效的统计资料,关系到统计能否得出正确的结论。从某种意义上讲,没有统计分组,就没有科学的统计资料的整理,也就没有科学的统计分析。统计分组绝不是一个单纯的技术问题,而是具有高度原则性和理论性的问题。理解统计分组的概念应注意三点:①统计分组的对象是总体,统计分组是把总体分为若干部分或组,例如,将 1000 家企业这一总体分为 5 组,将高校教师这一总体分为 5 组,都是将总体分为若干部分或组。②统计分组都应该有分组标志,分组标志是分组的依据,例如,将 1000 家企业按“所有制性质”分组,高校教师按“职称”分组等。当然,分组标志可以是品质标志,也可以是数量标志。③统计分组对总体而言是“分”,对总体单位而言是“合”,即统计分组把总体分为几个部分或组后,它同时也就把性质相同的一些总体单位组合在一起,因此统计分组与汇总是两个密切相关的统计资料整理方法。例如,把大学教师分为无职称、助教、讲师、副教授、教授 5 组,组与组之间教师的职称存在差别,同时相同职称的教师则被划归到一组,组合在一起了。

(2) 统计分组的作用。统计分组在统计研究中占有重要地位,其基本作用体现在以下几个方面。

① 划分现象的类型。社会经济等现象存在复杂多样的类型,各种不同的类型有不同的特点。在整理大量统计资料时,需要运用统计分组法将所研究的总体划分为不同的类型组来进行研究。例如,对某所大学教师按不同职称分组后,可分析不同职称教师人数、特点等。

② 揭示现象内部结构。社会经济等现象包括大量单位,不但在性质上不同,而且在总体中所占的比重也不一样。各组比重的大小不同,说明它们在总体中所处的地位不同,对总体特征的影响也不同。其中比重大的部分,很大程度上影响着总体的性质。例如,我国国内生产总值可分为第一产业、第二产业和第三产业,可以根据各产业的产值分析产业结构,进而分析国家经济的发展程度等。

如果将总体的结构分组资料按时间先后顺序排列,可以反映各个组成部分在总体中地位的变化情况,从而认识总体的发展变化规律。例如,从 2021—2025 年我国第一、第二、第三产业比重的变动情况来看,第一产业增加值占国内生产总值的比重基本逐年下降,第二产业增加值占国内生产总值的比重持续下降,第三产业增加值占国内生产总值的比重持续上升,具体如表 3-1 所示。

表 3-1 2021—2025 年第一、第二、第三产业增加值占国内生产总值的比重^① 单位：%

各产业增加值占国内生产总值的比重	2021 年	2022 年	2023 年	2024 年	2025 年
第一产业	7.1	7.1	6.9	6.8	6.7
第二产业	38.1	37.9	36.8	36.4	35.6
第三产业	54.8	55.0	56.3	56.8	57.7

③ 分析现象之间的依存关系。社会经济等现象之间存在着广泛的相互联系和制约关系,其中比较紧密的一种关系就是依存关系。研究现象依存关系的方法比较多,如相关与回归分析法、指数因素分析法、统计分组分析法等,其中统计分组分析法是最基本的方法,也是其他方法的基础。

统计分组分析法研究现象之间的依存关系,通常是把那些表现事物发展原因的因素称为影响因素,而把表现事物发展结果的因素称为结果因素。例如,收入和消费的关系,一般来讲,收入越高,消费也越多,这种依存关系称为正依存关系;职工家庭生活水平与家庭人口数之间的关系,一般来说,家庭人口数越多,在相同收入情况下,生活水平越低,这种依存关系称为负依存关系。

统计分组的上述三方面作用是分别从类型分组、结构分组和分析分组角度来说明的,它们不是彼此孤立的,而是相辅相成、相互补充、配合运用的。

(3) 统计分组方法。分组与汇总是统计整理的目的和核心,在设计调查表或调查问卷时就可以为分组与汇总做好基础工作,问卷中封闭式问题提供的答案就是分组。例如,性别,给出男性、女性的选择,就把被调查者按性别进行了分组,在整理时只需要汇总即可。当然也可以在汇总时同时完成分组工作,例如,汇总时把所有的被调查者的回答按男性、女性归类。

分组一般有三种方法:简单分组、平行分组、复合分组。①简单分组是将所有的总体按一个标志分组。例如,将所有被调查对象按性别分组,如表 3-2 所示。②平行分组是将所有的总体按两个及以上的标志平行分组。例如,按性别分成男性、女性,按年级分成大一、大二、大三、大四,如表 3-3 所示。③复合分组是将所有的总体按两个及以上的标志重叠分组。例如,先按性别分成男性、女性,然后男性及女性下面又按年级分成大一、大二、大三、大四,如表 3-4 所示。这种分组既可以得到一级分组中二级分组的结果,也可进行对比分析。例如,男生在大一至大四中的人数,大一中的男生和女生分别为多少? 一般简单分组极少用,因为调查时大部分要调查多个标志。至于具体如何分组,分组后如何汇总,将在本项目任务二编制分配数列中再详细介绍。

表 3-2 按性别分组

男	女

表 3-3 按性别、年级平行分组

按性别分		按年级分			
男	女	大一	大二	大三	大四

表 3-4 按性别、年级复合分组

男				女			
大一	大二	大三	大四	大一	大二	大三	大四

① 资料来源:国家统计局, https://www.stats.gov.cn/xxgk/sjfb/zxfb2020/202602/t20260228_1962662.html.



3. 绘制统计图表

整理后的统计资料,可以以统计表、统计图的形式表现出来。统计图表的绘制将在本项目的“任务三运用 Excel 进行统计整理”中介绍。

4. 统计资料的系统积累

有时因研究总体的需要,要将多次的统计资料整理后存档,并进行系统积累,以备统计分析时所需。例如,国民经济和社会发展统计公报中的部分内容包含历年的统计整理数据。

小 讨 论

如果要对某学院来自不同地区学生的月生活费收支情况进行分析,应该采用什么方式分组?某集团公司在全国有四个生产基地,同时生产 A、B、C、D、E 五种产品,如果要对四个生产基地生产的五种产品的销售利润率进行分析,应该采用什么方式分组?

三、统计资料整理方法

统计资料整理方法有手工和计算机两种方法。手工整理是指设计整理资料表格,根据审核过的统计表或调查问卷进行纸质稿的整理,适用于资料少、简单分析的统计资料。计算机整理是指在计算机中用统计分析软件(如 SPSS、Excel 等工具)整理统计数据或统计表、调查问卷等统计资料,适用于资料多、复杂分析的统计资料。无论采用哪种整理方法,在整理前应对已审核通过的统计表或调查问卷进行编号。

接下来,详细介绍在计算机中利用 Excel 进行资料整理的方法:①建立一个 Excel 文档。②设置行列的内容。第一列为资料(统计表或调查问卷)的编号,第一行为问题题号,封闭式问题的编号可设置成三种形式:第一种是答案唯一,一列为一题,如表 3-5 中的 1、2、3 题号;第二种是答案有多个但没有顺序要求,每个答案为 1 列,即一个题目多列,如表 3-5 中的多项选择题号 4.1~4.5;第三种是答案有多个且有顺序要求,即顺位式多项选择,如表 3-5 中的顺位式多选题号 5.1~5.5。③录入统计资料。按照问卷的编号将单项选择题、判断题的答案录入相应列的各行中;多项选择可以用 0 和 1 表示是否选择这个答案,如果没有选择这个答案录入 0,如果选择这个答案,则录入 1;顺位式多选题可以将排在第一位的答案编号放在该题的第一列,其他以此类推,如第三题为顺位式多项选择的结果 BDCAE,统计资料的具体录入样式可参考表 3-5 中的问卷编号 1 的录入结果,或者顺位式多选也可将答题者选择的顺序录入对应答案序号列中。开放式问题为了便于分析,可提炼共性的答案录入。例如,你认为当代大学生体能下降是因为什么导致的?答案有熬夜、晚睡、晚上玩游戏太晚等,其共性答案可以填不按时就寝。

表 3-5 利用 Excel 整理统计资料的样式

编号	题 号												
	1	2	3	4.1	4.2	4.3	4.4	4.5	5.1	5.2	5.3	5.4	5.5
1	A	C	D	1	0	1	1	0	B	D	C	A	E
2													

【任务实施】

- (1) 钱江集团人力资源部组织人员对员工满意度答题的 2000 份问卷进行审查,发现 42 份为重复 IP 提交的问卷、30 份问卷的答题时间不足 30 秒、18 份问卷出现逻辑矛盾或开放题中出现大量无意义字符,决定剔除这 90 份问卷,共得到有效问卷 1910 份。
- (2) 分组标准为工作 1~3 年、3~5 年、5~10 年以及 10 年以上,即按照工作年限划分为四组,并可以进一步根据员工性别(男、女)进行复合分组。详见表 3-6。

表 3-6 复合分组

男				女			
1~3 年	3~5 年	5~10 年	10 年以上	1~3 年	3~5 年	5~10 年	10 年以上
420	500	60	20	345	425	80	60

- (3) 因为采用问卷星调查,可以直接从问卷星平台导出 Excel 格式的答题结果,或者参照表 3-5 的样式整理 1910 份有效问卷。

【任务演练】

任 务 工 单

任务描述	各小组完成本小组回收调查问卷或调查表的审查、汇总
工具/资料/方法等	观看相关视频,学习“任务一初步整理统计资料”相关内容。如果通过问卷星开展调查问卷,请下载答题结果,完成审查
工作结果	回收调查问卷或调查表的整理结果
学生评(自评/互评)	
教师评(学校/企业)	

【演练评价】

评 价 标 准

学习目标	观察点	优秀 (100~90)	良好 (89~80)	中等 (79~70)	及格 (69~60)
能力目标	选择科学的方法进行审查。抽查审查结果,无任何问题	审查完成所有回收的调查问卷或调查表,且有问题的调查问卷或调查表都已审查出来	审查完成所有回收的调查问卷或调查表,但少量有问题的调查问卷或调查表未审查出来	审查完成所有回收的调查问卷或调查表,但有一半以上有问题的调查问卷或调查表未审查出来	审查完成所有回收的调查问卷或调查表,且有问题的调查问卷或调查表大部分未审查出来
素养目标	准时完成任务,遵守统计法律法规,结果真实、准确,整理方式方法有所创新,效率高,小组团结合作	准时完成任务,遵守统计法律法规,结果真实、准确,整理方式方法有所创新,效率高,小组团结合作	未达到优秀标准,其中有 1 个观察点未达到	未达到优秀标准,其中有 2 个观察点未达到	未达到优秀标准,其中有 3 个观察点未达到

未达及格及以上等级标准要求的为不及格,酌情赋分。如达到二维目标对应等级中的一项,则成绩降低一个等级。



任务二 编制分配数列

【任务描述】

某企业 50 名工人日加工零件的数据如下,要求采用组距式分组编制分配数列,并计算频数、频率、组中值。

520,520,520,520,550,550,580,580,580,580,540,540,540,540,540,540,540,
540,540,540,540,540,540,540,540,540,580,580,520,520,520,520,530,500,500,
500,500,500,510,510,520,520,520,500,510,510,500,500。

【任务准备】



微课：编制分配数列概述



微课：编制分配数列

一、分配数列的三个概念

要掌握分配数列,首先明确与分配数列有关的三个概念。

- (1) 分配数列。将总体按某一标志进行分组,并按一定顺序排列出每组的总体单位数,所得到的数列为分配数列。也就是统计整理的统计分组及计算分组后相关指标。
- (2) 频数。在分配数列中,分布在各组的总体单位数称为次数,又称频次。
- (3) 频率。各组次数与总次数之比称为比率,又称相对频率或相对次数。

二、分配数列的种类

分配数列按分配的标准不同分为品质数列和变量数列。品质数列是按品质标志分组形成的,用来观察总体单位中不同属性的单位分布情况的数列,如按性别分组。变量数列是按数量标志分组形成的,用来观察总体中不同变量值在各组的分布情况的数列,如按身高、体重分组。

三、分配数列的编制原则

品质数列的编制原则是只能是单项式分组,将品质性质相同的归为一组。例如,男性为一组,女性为一组。变量数列的编制原则分为连续变量和离散变量两种情况。连续变量只能按组距式分组,即以变量值的一定变动范围为一组,编制组距式变量数列(简称组距数列)。例如,按身高分为160~170厘米、170~180厘米等。离散变量一般按单项式分组,也可以是组距式分组,如果是单项式分组,则将每个变量值作为一组,编制单项式变量数列(简称单项数列)。例如,按班级学生年龄,19岁为一组、20岁为一组等。

无论采用哪种分组,分配数列编制结果一般是:组数不多不少、组内个体性质相同或接

近、每个个体可归入相应组,将其归纳为组数适当、组内同质、个体归组。

四、单项式分配数列的编制

品质标志只能是单项式分组,同一品质归为一组。例如,100 人分成男性 30 人,女性 70 人,其频数分别为 30 人和 70 人,频率分别为 30%和 70%,编制的分配数列以表格的形式呈现,具体如表 3-7 所示。数量标志中的离散变量一般进行单项式分组。例如,按年龄分成 19、20、21、22 四组,分别统计其频数和计算频率,编制的分配数列以表格的形式呈现,具体如表 3-8 所示。

表 3-7 品质数列的分配数列实例

性别	频数/人	频率
男	30	30%
女	70	70%
合计	100	100%

表 3-8 数量标志的单项式分配数列实例

年龄	频数/人	频率
19 岁	4	4%
20 岁	80	80%
21 岁	10	10%
22 岁	6	6%
合计	100	100%

五、组距式分配数列的编制

离散变量也可以进行组距式分组,连续变量只能是组距式分组,即编制组距式数列。编制组距式数列涉及以下几个要素:全距、组数、组距、组限、组中值、频数、频率、向上累计和向下累计。

1. 全距

如果总体变量有最大值与最小值,则全距=最大值-最小值。如果没有最大值、最小值,则无法计算全距,所编制的分配数列第一组和最后一组采用开口组,例如,50 以下、1000 以上。

2. 组数和组距

组数和组距之间可以互相转换,确定其中一个计算另一个。组距式变量数列,有等距和不等距两种。等距适用于标志值的变动比较均匀的情况和现象性质差异的变动比较均匀的情况。不等距适用于标志值的变动不均匀的情况和现象性质差异的变动不均匀的情况。

组数为总体分组的个数,组数=全距/组距,如果组数计算结果为小数,向上取整。

组距的确定分三种情况。①等距分组的组距=全距/组数;②不等距分组的组距=上限-下限;③如果组限不齐全,参考邻组的组距。

(1) 等距数列。等距数列是指每个组的组距都相等的组距数列。表 3-9 所示就是等距数列,各组的组距相等,均为 10。

等距数列有以下优点。

① 由于各组的组距相等,各组频数分布不受组距大小的影响,因而能直接比较各组频数的多少,便于反映总体单位的分布情况。

② 便于计算,尤其便于利用简捷法计算平均数、标准差等指标。所以,在一般情况下,应尽量采用等距分组来编制分布数列。

(2) 不等距数列。不等距数列又称异距数列,是指各组的组距不尽相等的组距数列。



如果变量值急剧地增长或下降,变动幅度很大,应采用异距分组来编制数列。例如,当前个体经营者的纯收入、农民的年收入等。还有,一些社会经济等现象,其性质的变化不是由变量值均匀变动而引起的。例如,为研究人口总体在人生各发展阶段的分布,就需要按照人在一生中自然的和社会的分组规律采用异距分组,编制异距数列。例如,按人口普查不同年龄分组有0~14岁、15~59岁、60岁及以上,如表3-10所示。这样各组的组距虽不相等,但能说明各自的问题。

表 3-9 某班学生统计学成绩

考分/分	人数/人	比重/%
50~60	2	5.0
60~70	7	17.5
70~80	11	27.5
80~90	12	30.0
90~100	8	20.0
合计	40	100.0

表 3-10 第七次全国人口普查全国人口年龄构成^①

人口年龄分组	人口数/人	比重/%
0~14岁	253383938	17.95
15~59岁	894376020	63.35
60岁及以上	264018766	18.70
合计	1411778724	100

异距数列各组频数的多少受组距大小的影响,不能直接准确反映实际的分布特征,各组频数也不能直接比较。为了消除组距不等的影响,必须计算频数密度或频率密度。计算公式为

$$\text{频数密度} = \frac{\text{某组的频数}}{\text{相应组的组距}} \quad \text{频率密度} = \frac{\text{某组的频率}}{\text{相应组的组距}}$$

3. 组限

组距式变量数列中,各组的界限称为组限。组限分为上限和下限,每组最大的标志值为上限,每组最小的标志值为下限。如果各组的组限齐全,称为闭口组。组限不齐全,缺上限或下限为开口组。

连续变量,采用“重叠分组、上限不在内”原则,每组变量值都以下限为起点,上限为极限,但不包括上限(最后一组上限除外)。离散变量,相邻组的上下限应当间断,但在实际中为求简便也可采用“重叠分组”。但要注意,当变量出现极大值或极小值时,可采用开口组,即用××以下或××以上表示,即为开口组。

一般来说,首先确定第一组的上下限,第一组的下限宜小于或等于最小的标志值,然后根据组距依次确定其他各组的组限,最后一组的上限宜大于或等于最大标志值。

4. 组中值

组距式分组,将一定范围的变量值归为一组,一般用组中值代表这组水平。闭口组的组中值=(上限+下限)/2。开口组的组中值计算只能参考邻组的组距,缺下限组的组中值=该组上限-邻组的组距/2;缺上限组的组中值=该组下限+邻组的组距/2。

5. 频数和频率

组距式分组中,分布在各组的总体单位数称为次数,又称频数。各组次数或频数与总次数或频数之比称为比率,又称频率。

^① 数据来源为第七次全国人口普查公报(第五号)——人口年龄构成情况。

组距式数列编制步骤可总结为五步：①数据从小到大排序；②计算全距；③确定组数或组距；④确定第一组的下限和上限，其他组限以此类推；⑤计算各组频数、频率，并计算组中值。

6. 向上累计和向下累计

可以根据各组频数和频率的计算结果再进行向上累计和向下累计，向上累计又称以下累计，或称较小制累计，是从变量值低的组开始，将各组频数或频率逐次向变量值高的组累计，说明某一组上限以下各组的累计频数或频率；向下累计又称以上累计，或称较大制累计，是从变量值高的组开始，将各组频数或频率逐次向变量值低的组累计，说明某一组下限以上各组的累计频数或频率。累计分布的特点是：最后一组的向上累计频数或第一组的向下累计频数等于总频数，最后一组的向上累计频率或第一组的向下累计频率等于 1（或 100%），累计频数分布是确定各种位置平均数的依据。

【任务实施】

根据任务描述的事项，采用以下步骤完成组距式分组。

（1）计算全距。某企业 50 名工人日加工零件的最大值为 580，最小值为 500，则全距＝最大值－最小值，即 $580-500=80$ 。

（2）确定组距或组数，推算出组数或组距。假设本例以组距为 20 分组，则组数＝全距/组距＝ $80/20=4$ （组）。

（3）确定每组的组限。确定第一组的下限，第一组的下限是关键一步，要求第一组下限要小于或等于最小标志值，然后再根据组距确定上限，例如，第一组下限为 500，上限＝下限＋组距＝ $500+20=520$ ，其他组限以此类推，各组组限如表 3-11 所示。

（4）计算各组的组中值。本例各组组限齐全，根据组中值＝（上限＋下限）/2 的式子计算各组的组中值，计算结果如表 3-11 所示。

（5）计算各组频数和频率。根据“重叠分组、上限不在内”原则计算各组的频数，具体如表 3-11 所示。然后根据频率＝频数/总体单位数×100%公式计算各组频率，计算结果如表 3-11 所示。

表 3-11 某企业 50 名工人日加工零件的数据整理成组距式分配数列结果

组限/个	频数/人	频率/%	向 上 累 计		向 下 累 计		组中值/个
			频数/人	频率/%	频数/人	频率/%	
500～520	12	24	12	24	50	100	510
520～540	12	24	24	48	38	76	530
540～560	20	40	44	88	26	52	550
560～580	6	12	50	100	6	12	570
合计	50	100					

（6）计算向上累计和向下累计。第一组向上累计频数为本组的频数 12，第二组向上累计频数为第一组频数 12 加上第二组频数 12 为 24，其他组的向上累计可将第一组频数与其他各组频数相加得到或用前一组的向上累计频数加上本组频数；依此类推可得到向上累计频率，如表 3-11 所示；第一组向下累计频数为频数的合计数 50，第二组向下累计频数为 50