



项目三 统计整理

学 而 思 行

所有的数据收集到以后是无法直接用于简单描述统计或计量分析的，这就需要我们具有较强的数据整理能力。企业组织利用相关数据进行分析可以帮助自身降低成本、提高效率、开发新产品、做出更明智的业务决策等。通过学习本章内容，学生将形成数据整理思维，利于有目标、有条理、高效地完成自己的工作。

知识目标

- 了解统计整理的概念和内容；
- 掌握统计分组的方法并在分组的基础上进行次数分配和数列的编制；
- 学会用统计表来表示统计资料。

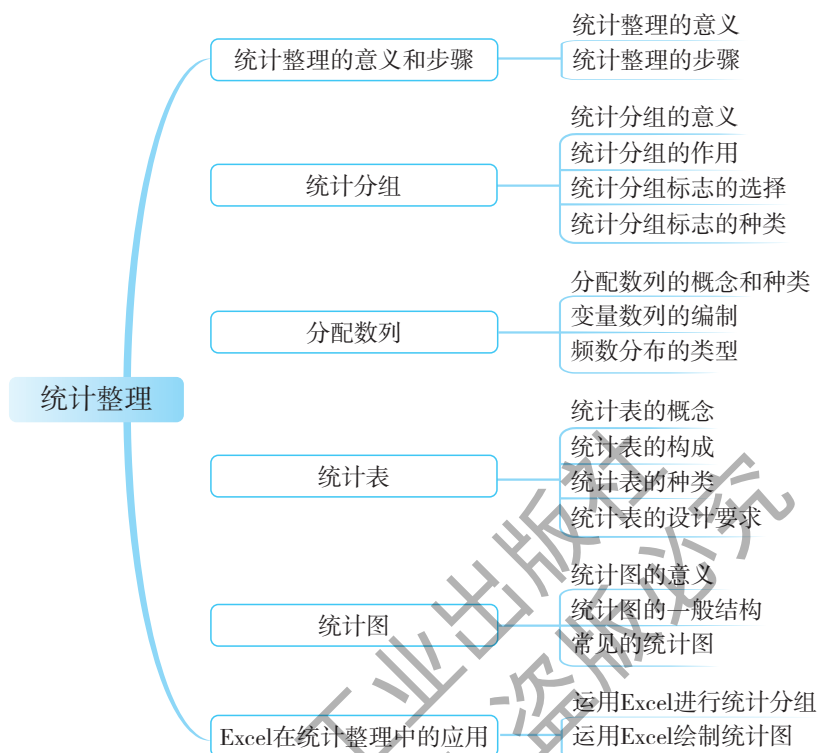
能力目标

- 能对调查资料按调查目的进行科学整理；
- 根据研究目的对统计资料进行有效分组，形成反映现象特征的分配数列；
- 能根据某一实际问题设计统计调查问卷；
- 熟练运用Excel进行统计分组。

素质目标

- 培养数据整理思维。

思维导图



情景引入

某县苗圃对 110 株树苗的高度进行测量(单位:厘米),数据如下:

154 133 116 128 85 100 105 150 118 97 110 131 119 103 93 108 100 111 130
 104 135 113 122 115 103 90 108 114 127 87 127 108 112 100 117 121 105 136
 123 108 89 94 139 82 113 110 109 118 115 126 106 108 115 133 114 119 104
 147 134 117 119 91 137 101 107 112 121 125 103 89 110 122 123 124 125 115
 113 128 85 113 143 80 102 132 96 129 83 142 112 120 107 108 111 100 97
 111 131 109 145 93 135 98 142 127 106 110 101 116 110 123

思考:

- (1) 就以上列举的 110 株树苗的高度数据看,分布呈现什么规律?
- (2) 怎样编制次数分布表?



任务一 统计整理的意义和步骤



一、统计整理的意义

统计人员通过统计调查,搜集到大量个体单位分散、零碎的原始资料,而统计研究的目的是要揭示社会经济现象总体的数量特征及其规律,统计整理的意义和步骤仅靠调查得来的资料难以达到这一目的,必须对这些资料加以整理、汇总,使之系统化、条理化。

统计整理是根据统计研究的目的和任务,将调查取得的大量原始资料进行科学地分类(或分组)、汇总,为统计分析提供条理化、系统化的综合统计资料的工作过程。有时,为了特定的目的,对已经整理过的统计资料(即次级资料)进行加工,以满足统计分析的要求,也属于统计整理工作的范围。

统计整理是一项非常重要的工作,既是统计调查的继续,又是统计分析的基础和前提条件,在整个统计研究中起着承前启后的作用。统计整理工作的质量直接影响到能否对社会经济现象进行准确的数量描述和数量分析。

二、统计整理的步骤

统计整理的目的是通过对大量原始资料进行加工整理,得到说明总体特征的综合资料,通过对事物个性的研究达到对事物共性的认识,从而揭示事物的发展规律。因此,统计整理是一项严密细致的、科学性很强的工作,需要有组织、有计划地进行,以下是统计整理的基本步骤。

(一) 设计统计整理方案

在进行统计整理之前,要根据统计研究的目的,确定统计指标及统计指标体系,确定统计分组的方法,确定统计汇总的方法,确定统计指标的表现形式等。统计整理方案是统计设计在统计整理阶段的具体化,是保证统计整理工作顺利进行的前提,因此务求详尽、具体。

(二) 对调查资料进行审核

为了确保统计工作的质量,在统计整理过程中首先要做好原始资料的审核工作。主要包括以下四点。

1. 准确性审核

准确性审核主要包括以下两方面:

(1) 计算检查。即审核资料的统计口径和范围、计算方法和计量单位等是否符合要求,计算结果是否准确,是否符合实际情况。

(2) 逻辑检查。即从理论上或根据常识来判断调查资料内容是否合情合理,各个项目之间是否有矛盾等。例如,在人口调查表中,在“与户主关系”栏填“父子”,而在“性别”栏却填“女”,其中必有一栏填错。

2.完整性审核

完整性审核就是审核调查中应调查的单位是否有遗漏或重复,调查表中应填的项目是否填写齐全等。如果调查单位不全、调查项目缺报,则据此整理出的资料就会不正确。

3.及时性审核

及时性审核就是审核资料是否按规定的时间上报,如果迟报,则需要分析迟报的原因。

4.适用性审核

适用性审核就是审核资料是否满足各地区、各部门的需要,是否具有现实意义。

(三)统计分组

只有按照最基本的、最能说明问题本质特征的统计分组和相应的统计指标对统计资料进行加工整理,才能对被研究的社会经济现象进行准确的数量描述和数量分析。因此,统计分组是统计整理的基础,分组科学与否直接影响到统计整理工作的质量。

(四)统计汇总

统计汇总是选择适当的汇总组织形式和技术方法,按分组要求对原始资料进行汇总,计算各组及总体的单位数和标志总量。统计汇总是统计整理的中心内容。

(五)编制统计表或绘制统计图

统计整理的结果需要用一定的方式表现出来,统计表和统计图是表现统计资料的两种主要方式,通过统计表或统计图表现统计资料,简洁明了,便于使用。

综上所述,设计整理方案、对原始资料进行审核是整理的前提,统计分组是统计整理的基础,统计汇总是统计整理的中心环节,编制统计表或绘制统计图是统计整理的结果。可见,统计整理主要包括统计分组、统计汇总和统计图表三方面的内容。



任务二 统计分组

一、统计分组的意义

(一)统计分组的概念

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的同质性。例如,社会产品按其经济用途分为生产资料和消费资料;工业企业按年产量或投资总额可分为大型、中型、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。

(二)统计分组的原则

统计分组是对总体各单位的分类,是整理统计资料的方法,也是统计分析的基础。因此,除了在内容上必须反映各单位、各组之间的性质差异外,还要在方法上保证资料的完整性和真实性,这是对统计分组最基本的要求。为此,统计分组在方法上必须符合以下两个原则。



统计分组

1. 穷尽性原则

穷尽性原则也叫不遗漏原则。即统计分组必须保证总体的每一个单位都能归入其中一个组,各个组的单位数之和等于总体单位总量,总体的指标必须是各个单位相应标志的综合。违背了这一原则,就会损害统计资料的完整性,从而也就损害了统计资料的真实性。

按照穷尽性原则分组时,需要重点注意的是分组的范围,它必须包括总体各单位在分组标志上的全部表现。即按品质标志分组时,组数是品质标志的全部类型;按变量分组时,最大组的上限应大于最大标志值,最小组的下限应小于最小标志值。

2. 互斥性原则

互斥性原则也叫不重复原则。即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,会影响到统计资料的真实性。在具体的分组过程中,为了保证各组之间不重复,按品质标志分组时,要重点注意对各组范围、特征、性质的界定,对于性质上较为复杂的单位要做出明确、统一的处理规定;按变量分组时,要重点注意相邻组之间重叠组限上的单位归属问题。统计的一般处理方法是重叠组限上的单位归入下限组,或者叫作“上限不在内”原则。同时还要注意,这种处理方法仅适合处理一般问题,而对于某些特殊问题则需要特殊处理。另外,统计上的这种一般处理与税法等其他学科的一般性处理也有所不同。

二、统计分组的作用

统计在认识社会经济现象的特征时,必须把总体数量特征与总体的结构、比例关系结合起来,才能全面、完整地揭示总体的数量特征。要实现这个目的,就需要进行统计分组。统计分组在统计研究中的作用主要体现在以下三个方面。

(一) 划分社会经济现象的类型

运用统计分组可以将复杂的社会经济现象划分成若干性质不同的类型,不同类型的社会经济现象可以反映出不同的特征。通过分组可以揭示不同类型的社会经济现象的本质和规律性。例如,我国的工业企业可划分为国有企业、集体企业、股份制企业、私营企业、外资企业等经济类型。

(二) 揭示社会经济现象总体的内部结构

对社会经济现象总体,按一定的标志进行分组,就可以研究总体的内部结构,而各部分所占比重的变化则可反映总体的性质、特点和发展变化规律。例如,某省 2016—2020 年国内生产总值构成变化情况如表 3-2-1 所示。

表 3-2-1 某省 2016—2020 年国内生产总值构成变化情况

| 年份 | 国内生产总值 | 第一产业/% | 第二产业/% | 第三产业/% |
|------|--------|--------|--------|--------|
| 2016 | 100 | 0.4 | 29.8 | 69.8 |
| 2017 | 100 | 0.4 | 30.5 | 69.1 |

续表

| 年份 | 国内生产总值 | 第一产业/% | 第二产业/% | 第三产业/% |
|------|--------|--------|--------|--------|
| 2018 | 100 | 0.3 | 29.8 | 69.9 |
| 2019 | 100 | 0.3 | 27.0 | 72.7 |
| 2020 | 100 | 0.3 | 26.6 | 73.1 |

从上表可以看出该省国民经济内部三次产业的构成及其变化情况,反映近几年我国产业结构的调整过程。

(三) 揭示社会经济现象之间的依存关系

社会经济现象不是孤立的,各种现象之间广泛存在着相互联系、相互依存、相互制约的依存关系。如自动化程度与劳动生产率之间、商品价格与市场供给量之间、商品销售额与流通费用率之间等,都在一定程度上存在相互依存的关系。当研究目的在于探讨同一总体范围内两个可变标志的依存关系时,可以将其中一个可变标志作为分组标志,以观察另一标志相应的变动情况。

例如,2014—2020年全国居民人均可支配收入与全国居民人均消费支出情况如表3-2-2所示。由表中数据可知,随着收入的增加,消费支出逐渐增加。

表 3-2-2 2014—2020年全国居民人均可支配收入与全国居民人均消费支出情况

单位:元

| 年份 | 可支配收入 | 消费支出 |
|------|----------|----------|
| 2014 | 20 167.1 | 14 491.4 |
| 2015 | 21 966.2 | 15 712.4 |
| 2016 | 23 821.0 | 17 110.7 |
| 2017 | 25 973.8 | 18 322.1 |
| 2018 | 28 228.0 | 19 853.1 |
| 2019 | 30 732.8 | 21 558.9 |
| 2020 | 32 188.8 | 21 209.9 |

从上表可以看出两列数据之间的依存关系。

以上统计分组的三方面作用是相互联系、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,它是一切统计工作的基础,应用于统计工作全过程,是统计研究的基本方法之一。

三、统计分组标志的选择

统计分组的关键是正确选择分组标志和划分各组的界限。分组标志的选择是否恰当将直接关系到统计分组是否正确,关系到能否实现统计研究的目的和任务。任何社会经济现象都有很多标志,要从许多标志中选择反映总体性质特征的标志,必须遵循以下三个基本原则。

(一) 根据研究的目的和任务选择分组标志

总体单位有许多标志,究竟选择什么标志作为分组标志,需要根据统计研究的目的和任务来决定。例如,在对某高校在校学生这一总体的研究中,每一个在校学生都是总体单位,学生有年龄、民族、性别、身高、学习成绩等许许多多的标志。而如果要了解学生的学习情况,就要选择学习成绩作为分组标志。

(二) 选择能反映现象本质的标志作为分组标志

在统计总体中,各单位具有许多标志,有些是主要的能反映事物本质特征的标志,有些则是次要的标志。例如,要研究企业经济效益的好坏,可供选择的标志很多,诸如工业产值、利税额、劳动生产率、资金利用率、人均利税额等,但最能综合反映企业经济效益好坏的标志是人均利税额。

(三) 根据现象所处的历史条件和经济条件选择分组标志

这是对第二点的补充和深化,因为在可供选择的许多标志中,衡量哪个是本质或主要的,往往离不开被研究现象所处的具体条件。也就是说,随着时间和具体条件的变迁,标志的重要程度可能会发生变化,如反映企业规模大小的标志有职工人数、营业收入、资产总额等,究竟应选择其中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,适宜用职工人数多少来表示企业规模大小;反之,在技术进步或技术装备比较先进的条件下,采用资产总额或营业收入就会更恰当,更切合实际。

四、统计分组标志的种类

根据统计研究的目的和任务,以及统计对象的特点不同,分组有以下几个基本类型。

(一) 按使用分组标志的特征不同,分为品质标志分组和数量标志分组

(1) 按品质标志分组,即按照事物属性的特征来分组。有比较简单的,如人口总体按性别分为男、女两组(2020年我国人口数及其构成如表3-2-3所示),民族按56个民族分组,这些组在性质上的差异比较明确和稳定。也有复杂的,如工业企业按产品形态和行业分组,人口按行业和职业分组,这些组的界限不易进行明确的划分。

表 3-2-3 2020 年我国人口数及其构成

| 按性别分组 | 人数/万人 | 比重/% |
|-------|---------|-------|
| 男 | 72 357 | 51.24 |
| 女 | 68 855 | 48.76 |
| 合计 | 141 212 | 100.0 |

(2)按数量标志分组,即按照反映事物差异的数量标志来分组。例如,人口总体按年龄分组;工业企业总体按产品产量分组;居民按收入水平分组等。按数量标志分组不是单纯确定各组的数量界限,而是通过数量差异区分各组的不同类型和性质。任何经济现象的质和量都是不可分割的,质规定着量,一定的量表现着一定的质。例如,2016—2020年全国居民的人均可支配收入按收入五等份分组情况如表3-2-4所示。

表 3-2-4 2016—2020 年全国居民的人均可支配收入按收入五等份分组情况

单位:元

| 组别 | 2016 年 | 2017 年 | 2018 年 | 2019 年 | 2020 年 |
|-------------------------|----------|----------|----------|----------|----------|
| 20%低收入组家庭 人均可支配收入 | 5 528.7 | 5 958.4 | 6 440.5 | 7 380.4 | 7 868.8 |
| 20%中间偏下收入组家庭 人均可支配收入 | 12 898.9 | 13 842.8 | 14 360.5 | 15 777.0 | 16 442.7 |
| 20%中间收入组家庭 人均可支配收入 | 20 924.4 | 22 495.3 | 23 188.9 | 25 034.7 | 26 248.9 |
| 20%中间偏上收入组家庭 人均可支配收入 | 31 990.4 | 34 546.8 | 36 471.4 | 39 230.5 | 41 171.7 |
| 20%高收入组家庭 人均可支配收入 | 59 259.5 | 64 934.0 | 70 639.5 | 76 400.7 | 80 293.8 |

注:全国居民按收入五等份分组是指,将所有调查户按人均可支配收入水平从低到高顺序排列,平均分为五个等份,处于最低20%的收入群体为低收入组,依次类推,为中间偏下收入组、中间收入组、中间偏上收入组、高收入组。本表数据为不同分组家庭的人均可支配收入。

(二)按照采用分组标志的多少不同,分为简单分组和复合分组

(1)简单分组:对研究总体按一个标志进行的分组。例如,将一个学校的学生按性别分组或按年龄分组就是简单分组,表3-2-3就是简单分组。简单分组操作容易,在现象总体结构简单的条件下,一般采用简单分组。

有时只凭一个标志进行一次分组是不够的,必须从不同的角度运用多个分组标志进行多方面的分组,才能达到统计分组的目的,此时,有必要对总体同时选择两个或两个以上的标志进行多次简单分组,这种多次简单分组称为平行分组或平行分组体系。例如,为认识运动员的自然构成,可以按照性别、年龄、身高、体重等标志进行多次分组。

①按性别分组:男、女。

②按身高分组(厘米):150以下,150~160,160~170,170~180,180以上。

③按年龄分组(周岁):8以下,8~12,12~14,14~16,16~18,18~25,25以上。

④按体重分组(千克):40以下,40~50,50~60,60~70,70~80,80以上。

平行分组的特点是每次分组只能固定一个因素对差异的影响,而不能固定其他因素对差