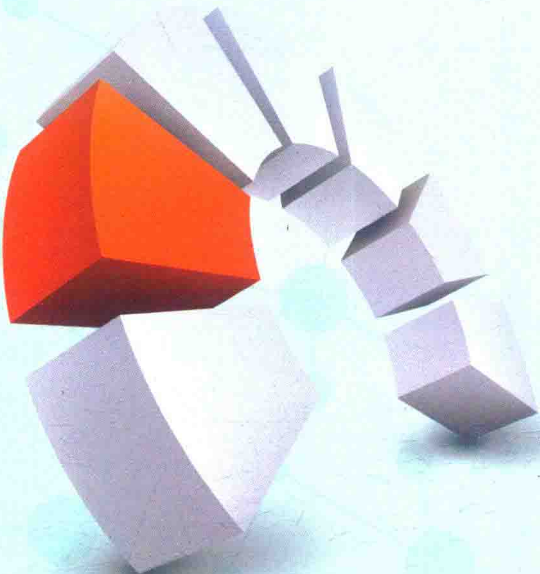
数理统计及实训教程

主编：杨晓刚　王　灿



华东师范大学出版社

数理统计及实训教程

主 编：杨晓刚　王　灿

华东师范大学出版社

图书在版编目(CIP)数据

数理统计及实训教程/杨晓刚,王灿主编. —上海:华东师范大学出版社,2018
ISBN 978-7-5675-7818-0

Ⅰ. ①数… Ⅱ. ①杨…②王… Ⅲ. ①数理统计—教材 Ⅳ. ①O212

中国版本图书馆 CIP 数据核字(2018)第 110607 号

数理统计及实训教程

主　　编　杨晓刚　王　灿
项目编辑　李　琴
特约审读　王小双
责任校对　朱　鑫
装帧设计　庄玉侠

出版发行　华东师范大学出版社
社　　址　上海市中山北路 3663 号　邮编 200062
网　　址　www.ecnupress.com.cn
电　　话　021-60821666　行政传真 021-62572105
客服电话　021-62865537　门市(邮购)电话 021-62869887
地　　址　上海市中山北路 3663 号华东师范大学校内先锋路口
网　　店　http://hdsdcbs.tmall.com

印 刷 者　浙江临安曙光印务有限公司
开　　本　787×1092　16 开
印　　张　10.5
字　　数　240 千字
版　　次　2018 年 7 月第 1 版
印　　次　2018 年 7 月第 1 次
书　　号　ISBN 978-7-5675-7818-0/G · 11178
定　　价　31.60 元

出 版 人　王　焰

本书编写委员会

主　编　杨晓刚　王　灿

编　委　王万禹　宿　娟　鄢盛勇　潘朝毅　李　虎
　　　　王成强　徐　祯　赵晓燕　饶若峰　谭启建

前言

数理统计乃数学中联系实际最直接最广泛的分支之一,但应该指出,其在研究方法上有它的特殊性,和其他数学学科的主要不同点有:

第一,由于随机现象的统计规律是一种集体规律,必须在大量同类随机现象中才能呈现出来,所以,观察、试验、调查就是概率统计这门学科研究方法的基石.但是,作为数学学科的一个分支,它依然具有本学科的定义、公理、定理.尽管这些定义、公理、定理来源于自然界的随机规律,但这些定义、公理、定理是确定的,不存在任何随机性.

第二,在研究概率统计中,使用的是“由部分推断全体”的统计推断方法.这是因为它研究的对象——随机现象的范围是很大的,在进行试验、观测的时候,不可能也不必要全部进行.但是由这一部分资料所得出的一些结论,要在全体范围内推断其可靠性.

第三,随机现象的随机性,是针对试验、调查之前而言的.而真正得出结果后,对于每一次试验,它只可能得到这些不确定结果中的某一种确定结果.我们在研究这一现象时,应当注意在试验前能不能对这一现象找出它本身的内在规律.

本书将着重介绍点估计(矩法估计、极大似然估计)、参数假设检验、非参数假设检验、回归分析和方差分析等基本知识和原理,并借助 SPSS 软件实现对应的实验,使读者对数理统计的原理和作用有更深刻的了解.希望本书可以帮助读者全面理解、掌握数理统计的思想与方法,既能掌握基本而常用的分析和计算方法,又能运用数理统计的观点和方法来研究解决经济与管理等统计工作中的实际问题.

编者

2018.05

目录

第一章

统计学基础知识

第一节 几个基本概念

一、统计总体和总体单位

统计总体简称总体，是指根据一定的研究目的，统计所要研究的、客观存在的、具有某一共同性质的许多个别单位所构成的整体. 构成总体的各个个别单位，就是总体单位，简称单位或个体，它是构成总体的最基本单位. 例如，要研究某市工业生产经营情况，该市所有的工业企业就是一个总体. 这是因为在性质上每个工业企业的经济职能是相同的，都是从事工业生产活动的基本单位，即它们是同性质的，而每一个工业企业就是一个总体单位.

统计总体根据总体单位是否可以计量分为有限总体和无限总体.

有限总体是指一个统计总体中包含的单位数是有限的. 例如，全国人口数、工业企业数、商店数等，不论它们的单位数量有多大，都是有限的，可以计量的. 对有限总体可以进行全面调查，也可以进行非全面调查.

无限总体是指一个统计总体中包含的单位数是无限的. 例如，工业生产中连续大量生产的产品、大海里的鱼资源数等，其数量都是无限的. 对无限总体不能进行全面调查，只能抽取一部分单位进行非全面调查，据此推断总体.

统计总体具有以下三个特征：

第一，同质性. 是指构成总体的各个单位必须具有某一个共同的特征和性质. 同质性是各个个别单位构成统计总体的先决条件.

第二，大量性. 是指总体是由许多单位组成的，仅个别或少数单位不能构成总体. 这是因为统计研究的目的是为了描述现象的规律，由于个别单位的现象有很大的偶然性，而大量单位的现象综合则相对稳定. 因此，现象的规律性只能在大量个别单位的汇总综合中才能表现出来.

第三，变异性. 是指构成总体的各单位只是在某一性质上相同，而在其他性质或特征上具有一定的差异. 例如，某市全体工业企业的经济职能相同，但是在所有制类型、经营规模、职工人数等方面是不同的. 同质性是构成总体的基础，变异性使统计研究成为必要. 如果总体的各个单位没有差异，统计研究就成了毫无意义的活动.

总体和总体单位具有相对性，它们随着研究目的不同是可以变换的. 例如，要研究某地区工业企业的生产经营情况，则该地区全部工业企业构成总体，而每一个工业企业是单

位:如果要研究该地区某一个企业的生产经营情况,那么该企业就成了总体,该企业下属的各个职能部门就是单位.由此可见,一个工业企业由于研究目的不同,既可以作为一个单位来研究,也可以作为一个总体进行研究.

二、指标与标志

(一)指标

指标,亦称统计指标,是说明总体现象数量特征的概念及其数值.统计指标有两种使用方法,一是进行统计设计或理论研究时所使用的仅有数量概念而没有具体数字的统计指标,例如国内生产总值、国民生产总值、商品销售额、人口出生率等.二是统计指标由指标名称和指标数值构成.例如,某年某市国内生产总值为 3000 亿元,它包括指标名称:国内生产总值;指标数值:3000 亿元.从完整的意义上讲,指标由六个要素构成:时间限制、空间限制、指标名称、指标数值、计量单位、计算方法.

统计学中通常把统计指标分为数量指标和质量指标.

数量指标是反映现象总规模、总水平和工作总量的统计指标,例如人口总数、职工总数、企业总数、工资总额、国内生产总值、商品销售额、货物运输量等.由于数量指标反映现象的总量,所以也称为总量指标,并且由于用绝对数表示,也称为统计绝对数.

质量指标是反映现象相对水平或工作质量的统计指标,例如人口密度、出生率、死亡率、出勤率、劳动生产率、单位产品成本、职工平均工资等.质量指标通常是由两个总量指标对比而派生的指标,用相对指标或平均指标来表示,反映现象之间的内在联系和对比关系.

(二)标志

标志是说明总体单位属性和特征的名称.例如,某企业全体职工作为一个总体,每一位职工是总体单位,职工的性别、年龄、籍贯、民族、文化程度、工龄、工资水平等是说明每一名职工的特征的名称,都称为标志.显然,总体单位是标志的承担者.

标志按其性质不同可分为品质标志和数量标志.

品质标志是表明总体单位品质属性或特征的名称,它不能用数值表示,只能用文字说明.例如,工业企业职工的性别、籍贯、民族、文化程度就是品质标志.

数量标志是表明总体单位数量特征的名称,是用数值表示的.例如,工业企业职工的年龄、工龄、工资水平就是数量标志.数量标志的具体表现为标志值,或为变量值.例如,某工业企业的某职工年龄为 38 岁,工资为 2800 元,其数值就是标志值.

(三)指标与标志的区别与联系

指标与标志既有明显的区别,又有密切的联系.两者的区别有以下两点:

(1)指标是说明总体特征的,而标志是说明总体单位特征的.

(2)标志有能用数值表示的数量标志和不能用数值标识的品质标志,而指标不论是数量指标还是质量指标,都是用数值表示的.

指标与标志的联系有以下两点:

(1)统计指标的数值是从总体单位数量标志的标志值进行直接汇总或间接计算的.

例如，某工业企业职工的月资总额是该企业的所属职工的月资额汇总而来的，而职工的月平均工资则是通过进一步计算得到的.

(2) 指标与数量标志之间存在着变换关系. 当研究目的发生变化，原来的统计总体如果变成了总体单位，则相对应的统计指标也就变为数量标志，反之亦然. 总之，统计指标与数量标志的变换关系和总体与总体单位的变换关系是一致的. 例如，研究目的由原来某地区工业企业的生产经营情况，变为只是研究该地区某一个工业企业的生产经营情况，那么该企业的工业增加值、职工人数、劳动生产率、工资总额等就由原来的数量标志变成为反映该工业企业总体特征的指标了.

三、变异、变量与变量值

变异是指统计所研究的指标与标志，其具体表现在总体及总体单位之间是可变的，即指标及标志的具体表现在各总体或各单位之间不尽相同或有差异. 这样的指标或为变异指标或变异标志. 变异指标是反映不同总体的同一指标之间数值的差异. 变异标志则是反映同一总体内同一标志不同单位之间的差异. 对于品质标志而言，是属性或特征的差异；对于数量标志而言，是数量上的差异. 变异是统计分组和统计分析的基础. 如果没有变异，也就没有必要进行统计研究了.

可变的统计指标和可变的数量标志称作变量. 变量是一种概念或名称，变量的具体数值或具体表现就是变量值，即变量值是指标数值或数量标志的标志值. 变量与变量值是两个既有密切联系又有明显区别的不同概念，不能混用. 例如，某车间有 4 名工人，其月产量分别为：1000 件、1200 件、1500 件、1800 件，这些都是“产量”这个变量的具体数值. 如果要计算这 4 名工人的月平均产量，不能说是求这 4 个变量的平均数，因为这里只有“产量”一个变量，并不是 4 个变量，而所要平均的是这一个变量的 4 个变量值.

变量按变量值是否连续可以分为连续变量和离散变量. 连续变量的变量值是连续不断的，相邻两个值之间可作无限分割，即可取无限多个数值. 例如，人的身高、体重、年龄、零件误差的大小等都是连续变量，它们可以通过称重、测量或计算取到小数后的任意一个位数. 而离散变量的变量值是有限个或可列无限多个，只能用计数的方法取得. 例如，人数、厂数、机器设备数等都可以用数字表示，它们都是离散变量.

变量按其性质不同可以分为确定性变量和随机性变量. 确定性变量是指影响变量值的变动，起某种决定性作用的因素，致使该变量值沿着一定的方向呈上升或下降的变动. 例如，随着人们生活水平的提高以及医疗卫生条件的完善这些确定性因素的影响，使人的期望寿命这个变量的变量值不断提高. 随机性变量是指变量值的变化受不确定因素的影响，变量值的变化没有一个确定的方向，有很大的偶然性. 例如，在同一台机器设备上加工某种机械零件，其尺寸大小总是存在差异. 造成这种差异的因素可能有：原材料质量的变化、电压的不稳定、气温和环境的变化以及操作工人的情绪波动等. 这些影响该种机械零件尺寸变动的因素都是随机发生的，是不确定的. 这里的机械零件尺寸就是一个随机性变量.

四、统计指标体系

统计指标体系是指由若干个相互联系的统计指标所构成的有机整体，用以说明所研

究的总体现象各方面的相互依存和相互制约的关系.

单个的统计指标只能反映总体现象的某一个侧面的特征,而一个总体往往具有多种数量表现和数量特征,并且彼此不是孤立的.如果要全面地认识总体的基本特征,必须将反映总体各方面特征的一系列统计指标结合起来,形成统计指标体系,使得我们对总体有更全面、更系统、更深入的认识,更好地发挥统计的整体功能.

由于总体现象本身的联系是多种多样的,所以统计指标之间的联系也是多种多样的,相应地可以建立各种各样的统计指标体系.例如,要反映工业企业的全面情况,就用一系列关于人力资源、资金、物资、生产技术、供应及销售等相互联系的指标来组成工业企业统计指标体系.如果只反映工业企业的产品生产量的情况,就可用产品实物量、产品品种、质里、总产值、净产值、原材料消耗、产品成本、销售利润等一系列统计指标构成产品生产量统计指标体系.如果要从宏观经济的角度反映国民经济运行不同环节之间的经济联系,就必须从生产、分配、流通、使用等过程相应地建立一系列指标,构建反映国民经济运行状况的统计指标体系.有些统计指标体系还可以用具体算术式表示,例如:

$$\text{商品销售额} = \text{商品价格} \times \text{商品销售量}.$$
$$\text{农作物收获量} = \text{亩产量} \times \text{播种面积}.$$

社会经济统计指标体系可以分为两大类:基本统计指标体系和专题统计指标体系.

基本统计指标体系是反映和研究国民经济与社会发展及其各个组成部分基本情况的指标体系,分为三个层次:最高层是反映整个国民经济与社会发展的统计指标体系,是由社会统计指标体系、经济统计指标体系、科技统计指标体系3个子系统构成;中间层则是各个地区和各个部门的统计指标体系,它是最高层统计指标体系的横向分支和纵向分支,是为了满足本地区和本部门的社会经济管理、检查、监督的需要而设置的指标体系;第三个层次是基层统计指标体系,是指各种企业和事业单位的统计指标体系.它既要满足本企业和本单位的管理和监督的需要,同时也要满足中间层和最高层建立统计指标体系的需要.

专题统计指标体系是针对社会经济的某一个专门问题而制定的统计指标体系.例如,经济效益指标体系、小康生活水平指标体系、和谐社会指标体系等.

统计指标体系按其功能不同,可分为描述统计指标体系、评价统计指标体系和预警统计指标体系.描述统计指标体系是全面反映客观事物的状况、运行过程和结果,它包括所有必要的统计指标,具有较强的稳定性.评价统计指标体系是比较、判断客观事物的运行过程和结果正常与否,它是根据不同分析评价的需要而建立的.它有一部分指标可以直接从描述统计指标体系中选取,另一部分指标可由描述统计指标加工处理后得到,该指标体系比较灵活、变动性大.预警统计指标体系是对客观事物的运行进行监测,并根据指标值的变化,预报即将出现的不正常状态、突发事件及某些结构性障碍等.该体系的指标一部分是由描述指标体系中的灵敏性和关键性指标所组成,另一部分是对一些描述指标加工而成.在这三种指标体系中,描述统计指标体系是最基本的指标体系,它是建立评价、预警统计指标体系的基础.

第二节　数据及其描述

一、统计数据的分类

按计量层次可以分为定类数据、定序数据、定距数据、定比数据.

二、数值平均数

数值平均数是对统计数列的所有各项数据计算的平均数，它能够概括整个数列中所有各项数据的一般水平和集中趋势，并受数列中每一个标志值变动的影响. 数值平均数主要有：算术平均数、调和平均数和几何平均数.

(一) 算术平均数($\bar{x}$)

算术平均数是一种运用最广泛、最频繁的平均数，它是将总体各单位某一数量标志之和求得标志总量后，除以总体单位总数. 当提到平均数而又未说明其形式时，通常就指算术平均数，其基本公式如下：

$$\text{算术平均数}=\frac{\text{总体标志总量}}{\text{总体单位总数}}.$$

利用这一计算公式时，应注意公式的分子项与分母项在总体范围上必须保持一致，否则，其意义与平均指标就有所不同. 这也是平均指标与相对指标的性质差异.

根据所掌握的资料不同，算术平均数可以分为简单算术平均数和加权算术平均数. 在具体的计算过程中，根据未经分组的原始数据求平均时，一般计算简单算术平均数，且简单算术平均数多用于数据量较小的情况. 当数据量较大时，我们一般用分组或频率分布进行计算——以频率或频数为权数，用加权平均数的形式计算算术平均数.

1. 简单算术平均数

简单算术平均数就是直接将各变量值相加，再除以变量值的个数. 简单算术平均数在资料未经分组整理的情况下应用，其计算公式为：

$$\bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{\sum x}{n}.$$

其中，$\bar{x}$表示算数平均数；x_i 表示第 i 个单位的标志值 ($i=1,2,3,\cdots,n$)；n 表示总体单位总数.

例 1. 某生产小组有 10 名工人，日生产零件分别为 34 件、28 件、35 件、45 件、42 件、37 件、30 件、40 件、38 件、43 件，求该 10 名工人的人均日产量.

解：10 名工人的人均日产量为：

$$\bar{x}=\frac{\sum x}{n}=\frac{34+28+35+45+42+37+30+40+38+43}{10}=37.2(\text{件 / 人}).$$

2. 加权算术平均数

当资料已经分组，整理成变量数列时，可以使用加权平均数来计算. 其计算公式为：

$$\bar{x}=\frac{x_1f_1+x_2f_2+\cdots+x_nf_n}{f_1+f_2+\cdots+f_n}=\frac{\sum_{i=1}^{n}x_if_i}{\sum_{i=1}^{n}f_i}=\frac{\sum xf}{\sum f}.$$

其中，x_i 表示第 i 组变量值；f_i 表示第 i 组单位数；x_if_i 表示第 i 组的标志量总和（$i=1, 2, \cdots, n$）；n 表示组数.

上式也可以用公式表示为：

$$\begin{aligned}\bar{x}&=\frac{\sum xf}{\sum f}=\frac{x_1f_1+x_2f_2+\cdots+x_nf_n}{\sum f}\\&=x_1\frac{f_1}{\sum f}+x_2\frac{f_2}{\sum f}+\cdots+x_n\frac{f_n}{\sum f}\\&=\sum x\frac{f}{\sum f}.\end{aligned}$$

例2. 某工厂车间20名工人加工某种零件的日产量资料如表1-1所示，试计算这20名工人的平均日产量.

表1-1　20名工人零件生产数量分组资料

按日产量分组（件）	工人人数（人）
14	2
15	4
16	8
17	5
18	1
合计	20

解： 平均日产量计算表如表1-2所示.

表1-2　20名工人平均日产量计算表

按日产量分组（件）x	工人人数（人）f	总产量（件）xf	各组工人人数占总人数比重 $\frac{f}{\sum f}$	$x\frac{f}{\sum f}$
14	2	28	0.10	1.40
15	4	60	0.20	3.00
16	8	128	0.40	6.40
17	5	85	0.25	4.25
18	1	18	0.05	0.90
合计	20	319	1.00	15.95

20 名工人平均的日产量为：

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{14\times2+15\times4+16\times8+17\times5+18\times1}{2+4+8+5+1}=\frac{319}{20}=15.95(\text{件}/\text{人}).$$

如果利用工人比重的资料进行加权计算，也可以得到同样的结果：

$$\begin{aligned}\bar{x}&=\sum x\frac{f}{\sum f}=14\times0.1+15\times0.2+16\times0.4+17\times0.25+18\times0.05\\&=15.95(\text{件}/\text{人}).\end{aligned}$$

根据数据资料的不同，用来作为权数的主要有两种形式：一种是数据的各可能值——变量出现的次数（频数），另一种是频率. 通过上面的举例可以看出，在计算加权算术平均数的过程中，无论是采用绝对权数（例中为工人人数），还是采用相对权数（例中为工人比重）来加权，其计算结果是一样的. 但从分析的角度来说，两种加权方式各有特点. 采用绝对权数计算，能够分别给出总体的单位总数（员工总人数）和标志总量（工资总额）；采用相对权数计算，则更能体现加权作用的实质. 这是因为：在被平均变量的可能取值已经给定的情况下，绝对权数的变化不一定会引起平均数计算结果的变化；而相对权数一旦变化，就必然会影响到平均数的计算结果. 所以，利用相对权数来分析总体内部结构变化对于平均数变化的影响，无疑具有独特的作用. 当然，当各变量值的权数都相等时，即 $f_1=f_2=\cdots=f_n$ 时，权数也就失去了衡量轻重的作用，这时加权算术平均数即为简单算术平均数.

3. 算术平均数的数学性质

算术平均数是最重要的平均数形式，了解和运用一些算术平均数的计算或分析性质，能够帮助我们在计算中减少工作量. 归纳起来，算术平均数有两个重要的数学性质.

(1) 算术平均数与各个变量值的离差之和为零.

对于简单算术平均数：$\sum(x-\bar{x})=0$ 或 $n\cdot\bar{x}=\sum x$.

证明：$\sum(x-\bar{x})=\sum x-n\cdot\bar{x}=\sum x-n\dfrac{\sum x}{n}=\sum x-\sum x=0$.

对于加权算术平均数，则有：$\sum(f\bar{x})=\sum fx$ 或 $\sum(x-\bar{x})f=0$.

证明：

$$\begin{aligned}\sum(x-\bar{x})\cdot f&=\sum xf-\sum\bar{x}f=\sum xf-\bar{x}\sum f\\&=\sum xf-\frac{\sum xf}{\sum f}\sum f=\sum xf-\sum xf=0.\end{aligned}$$

这些都表明，算术平均数用来代表个别单位的标志值固然存在误差，但用来代表整个总体或分布数列的一般水平，却是没有误差的，因为它与个别单位标志值的正、负离差恰好相互抵消，从而使得离差总和恒等于零. 这个性质说明，平均数是把总体单位变量值的差异全部抽象化了.

(2) 给定任意一个常数 c，对于简单算术平均数和加权算术平均数，分别有：

$$\sum(x-\bar{x})^2\leqslant\sum(x-c)^2,\ \sum(x-\bar{x})^2f\leqslant\sum(x-c)^2f.$$

证明：设 x_0 为不等于平均数$\bar{x}$的任意值，则会 $\bar{x}-x_0=c,\ c\neq 0$.

由于 $\bar{x}-x_0=c$，则 $x_0=\bar{x}-c$，代入以 x_0 为中心的离差平方和，得

$$\begin{aligned}\sum(x-x_0)^2&=\sum[x-(\bar{x}-c)]^2\\&=\sum(x-\bar{x}+c)^2\\&=\sum[(x-\bar{x})^2+2c(x-\bar{x})+c^2]\\&=\sum(x-\bar{x})^2+2c\sum(x-\bar{x})+nc^2\\&=\sum(x-\bar{x})^2+nc^2,\end{aligned}$$

从而有 $\sum(x-x_0)^2-nc^2=\sum(x-\bar{x})^2$.

由于 $c\neq 0$，则 $nc^2\geqslant 0$，

得 $\sum(x-x_0)^2\geqslant\sum(x-\bar{x})^2$，

故 $\sum(x-\bar{x})^2\leqslant\sum(x-c)^2$.

(二) 调和平均数(H)

调和平均数也称“倒数平均数”，它是对变量值的倒数求平均值，然后再取倒数而得到的平均数，记作 H. 作为算术平均数的一种变形，一种特定意义上的调和平均数，在统计中具有相当强的实用性. 调和平均数有简单调和平均数与加权调和平均数两种计算形式：

1. 简单调和平均数

$$H=\frac{1}{\dfrac{\dfrac{1}{x_1}+\dfrac{1}{x_2}+\cdots+\dfrac{1}{x_n}}{n}}=\frac{n}{\dfrac{1}{x_1}+\dfrac{1}{x_2}+\cdots+\dfrac{1}{x_n}}=\frac{n}{\sum\dfrac{1}{x}}.$$

2. 加权调和平均数

$$H=\frac{1}{\dfrac{\dfrac{m_1}{x_1}+\dfrac{m_2}{x_2}+\cdots+\dfrac{m_n}{x_n}}{m_1+m_2+\cdots+m_n}}=\frac{m_1+m_2+\cdots+m_n}{\dfrac{m_1}{x_1}+\dfrac{m_2}{x_2}+\cdots+\dfrac{m_n}{x_n}}=\frac{\sum m}{\sum\dfrac{m}{x}}.$$

例3. 某企业分三批购进同一种原材料，已知每批原材料购进的价格与购进的总金额如表 1-3 所示，试计算购进该种原材料的平均价格.

表 1-3　　购进原材料平均价格计算表

购进批次	价格(元/公斤)x	金额(元)m	购进数量(公斤) $f=\dfrac{m}{x}$
第一批	80	40 000	500
第二批	85	38 250	450
第三批	78	46 800	600
合计	—	125 050	1550

解:原材料平均价格为:

$$H=\frac{\sum m}{\sum \frac{m}{x}}=\frac{40\,000+38\,250+46\,800}{\frac{40\,000}{80}+\frac{38\,250}{85}+\frac{46\,800}{78}}=\frac{125\,050}{1550}=80.68(\text{元}/\text{公斤}).$$

该例题中:原材料平均价格是总金额(总体标志总量)除以购进总量(总体单位总量),它的计算方法实际上与算术平均数一样.调和平均数的权数即购进金额为购进价格与购进数量的乘积,即 $m=xf$,调和平均数和算术平均数的关系如下:

$$H=\frac{\sum m}{\sum \frac{m}{x}}=\frac{\sum xf}{\sum \frac{xf}{x}}=\frac{\sum xf}{\sum f}=\bar{x}.$$

本例中,利用加权算术平均数的式子可得

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{80\times 500+85\times 450+78\times 600}{500+450+600}=\frac{125\,050}{1550}=80.68(\text{元}/\text{公斤}).$$

由以上计算结果可见,调和平均数是算术平均数的变形,虽然它们的计算方法不同,但其实质是一样的.

对于调和平均数和算术平均数,若已知条件为分组资料的各组变量值 x,即各组的标志值总和 m 即 xf 时,可采用加权调和平均法计算平均指标;若已知条件为分组资料的各组变量值 x 即各组的次数 f 时,可直接用加权算术平均方法计算平均指标.

例4. 某年某集团公司下属有15个企业,其工业生产产值的计划完成程度即实际产值如表1-4所示,试计算该集团公司该年的平均完成程度.

表1-4　　购进原材料的平均价格计算表

产值计划完成程度(%)	组中值(%)x	企业个数	实际产值(万元)m	计划产值(万元)$\frac{m}{x}$
90～100	95	2	760	800
100～110	105	6	3675	3500
110～120	115	4	4485	3900
120～130	125	3	6000	4800
合计	—	15	14 920	13 000

解:根据各组已知的实际产值除以产值计划完成程度可求得各组计划产值,该集团公司的平均计划完成程度为:

$$\text{平均计划完成程度}=\frac{\text{实际产值总数}}{\text{计划产值总数}}\times 100\%=\frac{\sum m}{\sum \frac{m}{x}}\times 100\%$$

$$=\frac{760+3675+4485+6000}{\frac{760}{0.95}+\frac{3675}{1.05}+\frac{4485}{1.25}+\frac{6000}{1.25}}\times 100\%=\frac{14\,920}{13\,000}\times 100\%$$
$$=114.77\%.$$

(三) 几何平均数(G)

几何平均数是若干变量值的连乘积的 n 次方根,其中 n 是变量值的个数,几何平均数说明事物在一段时间按几何级数规律变化的量的平均水平,它主要用来计算平均发展速度. 几何平均数记作 G,根据掌握的资料是否分组,几何平均数也分为简单几何平均数与加权几何平均数两种方法.

1. 简单几何平均数

$$G=\sqrt[n]{x_1x_2\cdots x_n}=\sqrt[n]{\prod_{i=1}^{n}x_i}=\sqrt[n]{\prod x}.$$

其中,x_i 表示被平均的变量,$i=1,\ 2,\ 3,\ \cdots,\ n$;$\prod$ 表示连乘符号.

例 5. 某产品经过三个流水连续作业的车间加工生产而成. 本月第一车间的产品合格率为 90%,第二车间的产品合格率为 80%,第三车间的产品合格率为 70%. 求全厂的平均合格率.

解: 全厂的总合格率为:

$$总合格率=90\%\times 80\%\times 70\%=50.4\%.$$

因此平均合格率为:

$$平均合格率=\sqrt[n]{\prod x}=\sqrt[3]{90\%\times 80\%\times 70\%}=\sqrt[3]{50.4\%}=79.58\%.$$

2. 加权几何平均数

$$G=\sqrt[\sum f]{x_1^{f_1}x_2^{f_2}\cdots x_n^{f_n}}=\sqrt[\sum f]{\prod x^f}.$$

其中,f_i 表示各个变量值出现的次数,$i=1,\ 2,\ 3,\ \cdots,\ n$.

例 6. 设某笔为期 20 年的投资按复利计算收益,前 10 年的年利率为 10%,中间 5 年的利率为 8%,最后 5 年的年利率为 6%. 求年平均利率.

解: 年平均本利率 $=\sqrt[20]{1.1^{10}\times 1.08^5\times 1.06^5}=108.49\%$,

年平均利率 $=108.49\%-1=8.49\%$.

3. 几何平均数的数学性质

以 G 表示几何平均数,则几何平均数具有如下性质:

(1) $\ln G$ 等于 $\ln x_1,\ \ln x_2,\ \cdots,\ \ln x_n$ 的算术平均数.

(2) 设 G_x 和 G_y 分别为 $x_1,\ x_2,\ \cdots,\ x_n$ 和 $y_1,\ y_2,\ \cdots,\ y_n$ 的几何平均数,则 $x_1y_1,\ x_2y_2,\ \cdots,\ x_ny_n$ 的几何平均数等于 G_xG_y,$x_1/y_1,\ x_2/y_2,\ \cdots,\ x_n/y_n$ 的几何平均数等于 G_x/G_y.