

巍巍交大 百年书香  
www.jiaodapress.com.cn  
bookinfo@sjtu.edu.cn



策划编辑 袁相芬  
责任编辑 胡思佳 柳卫清  
封面设计 张瑞阳

# 统计学基础

TONGJIXUE JICHU



扫描二维码  
关注上海交通大学出版社  
官方微信



定价: 39.00元

中等职业学校财经商贸系列教材  
中等职业教育新形态一体化教材

统计学基础

主编 马瑞山



上海交通大学出版社

中等职业学校财经商贸系列教材  
中等职业教育新形态一体化教材

# 统计学基础

TONGJIXUE JICHU

主编 马瑞山



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

免费提供

精品教学资料包

服务热线: 400-615-1233  
www.huatengzy.com

中等职业学校财经商贸系列教材  
中等职业教育新形态一体化教材

# 统计学基础

TONGJIXUE JICHU

主编 马瑞山



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

## 内容提要

本书共包括七个项目,包括认识统计、了解统计调查、熟悉统计整理、掌握统计综合指标、掌握抽样技术、熟悉时间序列分析、了解统计指数。

本书可作为中等职业教育物流、市场营销、电子商务、旅游管理、财务会计、金融保险和其他经管类专业基础课程的教材,也可供相关企业人员参考。

## 图书在版编目(CIP)数据

统计学基础/马瑞山主编. —上海:上海交通大学出版社,2023. 1

ISBN 978-7-313-27316-1

I. ①统… II. ①马… III. ①统计学—中等专业学校—教材 IV. ①C8

中国国家版本馆 CIP 数据核字(2023)第 004494 号

## 统计学基础

TONGJIXUE JICHU

主 编:马瑞山

出版发行:上海交通大学出版社

邮政编码:200030

印 制:三河市骏杰印刷有限公司

开 本:787 mm×1 092 mm 1/16

字 数:186 千字

版 次:2023 年 1 月第 1 版

书 号:ISBN 978-7-313-27316-1

定 价:39.00 元

地 址:上海市番禺路 951 号

电 话:021-64071208

经 销:全国新华书店

印 张:13

印 次:2023 年 1 月第 1 次印刷

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系

联系电话:0316-3662258



# 前言 PREFACE

随着经济的高速发展，社会对财会从业人员的要求越来越高。培养高素质的会计从业人员是中等职业学校职业教育的重要任务之一。

考虑到中职学生的就业需求，我们对本书的教学内容进行了全方位的筛选。本书以培养学生岗位职业能力为核心，以工作过程为导向，以“教、学、做”一体化为目标，运用通俗的语言对统计学基础知识进行讲解。

本书在以下几个方面有所突破。

（1）体现课程思政建设的理念。为推进习近平新时代中国特色社会主义思想进教材进课程进头脑，本书在编写过程中融入课程思政的相关内容，不仅可以使学生掌握统计学专业专业知识，丰富学识，增长见识，塑造品格，而且可以增强学生的民族自豪感，激发学生的爱国情怀。

（2）内容突出对学生职业技能的训练。本书特别设置了“拓展窗”“想一想”等栏目，理论知识的选取紧紧围绕项目任务完成的需要进行，做到“适用、够用”。

（3）打破传统的统计学基础教材结构体系，采取项目引导、任务驱动的模式，按照知识点的难易程度和学生接受程度依次设置统计学基础项目。

（4）结构上，各项目前设有知识目标和能力目标，为学生指引学习方向；项目后有“职业基础知识测试”“职业核心能力训练”，有利于培养和提高学生的自主学习能力，提高其综合素质和实践技能。





本书建议 30 学时，实际学时可依据学校情况灵活调整，各项目的学时分配如下表：

| 内 容          | 学 时 |
|--------------|-----|
| 项目一 认识统计     | 3   |
| 项目二 了解统计调查   | 5   |
| 项目三 熟悉统计整理   | 4   |
| 项目四 掌握统计综合指标 | 5   |
| 项目五 掌握抽样技术   | 4   |
| 项目六 熟悉时间序列分析 | 5   |
| 项目七 了解统计指数   | 4   |
| 合计           | 30  |

本书由甘肃能源化工职业学院马瑞山副教授任主编。本书在编写过程中，得到了诸多从事统计学教学和研究工作的同仁的帮助，在此一并表示衷心的感谢！

由于编者水平有限，书中存在的不足之处，欢迎广大读者提出宝贵意见，以便我们进一步修改和完善。

编 者

# 目录

## CONTENTS

### 认识统计 项目一

- 任务一 初识统计学 ..... 3
- 任务二 掌握统计学中常用的基本概念 ..... 15
  - 【职业基础知识测试】 ..... 23
  - 【职业核心能力训练】 ..... 24

### 了解统计调查 项目二

- 任务一 认识统计调查 ..... 28
- 任务二 掌握统计调查方法 ..... 32
- 任务三 设计统计调查方案 ..... 37
  - 【职业基础知识测试】 ..... 40
  - 【职业核心能力训练】 ..... 41

### 熟悉统计整理 项目三

- 任务一 认识统计整理 ..... 45
- 任务二 理解统计分组 ..... 48
- 任务三 掌握频数分布 ..... 55
- 任务四 了解统计汇总与显示 ..... 61
  - 【职业基础知识测试】 ..... 77
  - 【职业核心能力训练】 ..... 77

### 掌握统计综合 指标 项目四

- 任务一 认识总量指标 ..... 81
- 任务二 掌握相对指标 ..... 86
- 任务三 掌握平均指标与标志变异指标 ..... 92
  - 【职业基础知识测试】 ..... 107
  - 【职业核心能力训练】 ..... 107



掌握抽样技术  
项目五

|                  |     |
|------------------|-----|
| 任务一 认识抽样调查 ..... | 111 |
| 任务二 掌握抽样误差 ..... | 121 |
| 任务三 理解参数估计 ..... | 127 |
| 【职业基础知识测试】 ..... | 134 |
| 【职业核心能力训练】 ..... | 135 |

熟悉时间序列  
分析  
项目六

|                           |     |
|---------------------------|-----|
| 任务一 认识时间序列 .....          | 139 |
| 任务二 掌握时间序列的水平分析 .....     | 144 |
| 任务三 掌握时间序列的速度分析 .....     | 153 |
| 任务四 了解长期趋势分析和季节变动分析 ..... | 159 |
| 【职业基础知识测试】 .....          | 169 |
| 【职业核心能力训练】 .....          | 170 |

了解统计指数  
项目七

|                  |     |
|------------------|-----|
| 任务一 认识统计指数 ..... | 173 |
| 任务二 掌握综合指数 ..... | 177 |
| 任务三 了解指数体系 ..... | 186 |
| 任务四 掌握平均指数 ..... | 194 |
| 【职业基础知识测试】 ..... | 197 |
| 【职业核心能力训练】 ..... | 198 |

|                  |     |
|------------------|-----|
| 附录 正态分布概率表 ..... | 200 |
| 参考文献 .....       | 202 |



# 项目一 认识统计

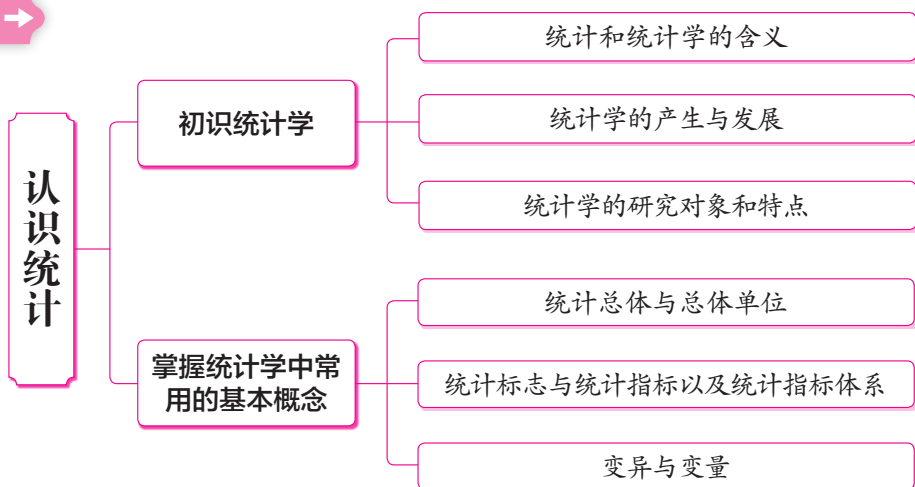
## 知识目标 →

- 了解统计学的研究对象并掌握其特点
- 了解统计的组织和管理

## 能力目标 →

- 学会运用统计研究的具体方法
- 能够利用统计方法完成统计工作

## 思维导图 →





## 项目情境 →

随着年龄的增长，近视的学生人数越来越多。学生的视力情况呈下降趋势，影响视力的主要原因是长时间用眼不当、眼睛疲劳、长时间看手机或电脑、写字和看书等姿势不良。作为某中职学校校长的王先生，想了解本学校学生的视力情况，于是对全校学生进行调查，分别统计出各个年级男生和女生近视的人数。通过调查发现：全校共有 2 000 名学生，近视学生的人数有 800 人，其中男生 450 人，女生 350 人。由此王先生可以清楚地了解到全校学生的视力情况。

通过上述内容，你还能想到生活中有哪些运用统计学知识的事例？

---

---

---

---

---

---

---

---

---

---



## 任务一 初识统计学

### 任务卡 1

箱子中有 10 个小球，白色的有 2 个，红色的有 8 个，现在从中取出 1 个小球，小张说抓到红色的概率比较大。上述内容有没有运用到统计学知识呢？

### 知识卡 1 统计和统计学的含义

在日常生活中，人们离不开统计数据。例如，老师每天上课前都要统计出勤人数，学生考试后要统计总成绩、平均成绩、及格率和优良率等，企业管理人员要统计供产销和盈利等，这些数据就是统计数据。统计数据是人们通过实际统计活动获得的，统计就是一门研究数据的技术。

统计是人们正确运用统计理论和方法采集数据、整理数据、分析数据和由数据得出结论的实际操作过程，是人们从数据上对客观世界的一种认识活动和结果。因此，统计活动的中心问题就是要获取数据与得出结论，进而向人们提供服务。

现代统计的含义包括三个方面：统计工作（统计活动）、统计资料和统计学。统计资料是统计工作的成果，统计学是对统计工作的经验总结和理论概括，反过来又指导统计工作，使统计资料准确可靠。

统计学是一门方法论学科，研究如何收集、整理、显示和分析统计数据，以探索现象内在的数量特征和数量规律。

统计学是和社会实践活动紧密关联的一门工具性学科，它产生于人类社会





实践的需要，并伴随着社会的发展、人的认识能力的提高而不断发展、成熟；反过来，它又为人类的社会实践活动提供强有力的支持。在人类的社会实践活动中，不论是宏观国民经济，还是在人们的日常生活和工作中，都存在大量的统计数据。

### 任务卡 2

某集团准备在某地新建一家零售商店，命令企划部经理高先生做好商店选址的准备工作。高先生明白，经过该地的行人数量是要重点统计的对象。于是，他委托相关人员进行了两个星期的观察，得到每天经过该地的人数。具体如下：

544, 468, 399, 759, 526, 212, 256, 456, 553, 259, 469, 366, 197, 178

如果设立商店要求的最低行人数量为 520 人，则根据上述观察数据，能否支持设店的决策？其中蕴含了哪些统计学知识？

## 知识卡 2 统计学的产生与发展

无论是在中国还是在外国，历朝历代、各个政府都会积极地利用统计活动为国家管理服务。

### 1. 古典统计学的产生与发展

古典统计学按照时间顺序可分为国势学派和政治算术学派。

(1) 国势学派产生于 18 世纪的德国，创始人是赫尔曼·康令和高尔费里德·阿亨华尔。1660 年，康令在大学里开设了国势学课程，并把国势学从法学、史学和地理学等学科中独立出来。1749 年，阿亨华尔在《近代欧洲各国国家学纲要》一书中首先使用了“统计学”一词，他收集了大量的实际资料，分门别类地记述了有关国情、国力的系统知识，进一步发展了国势学派的统计知识



体系。

(2) 政治算术学派产生于 17 世纪中叶的英国, 主要代表人物是威廉·配第和约翰·格朗特。英国古典政治经济学的创始人威廉·配第博士的代表作是《政治算术》, 书中提出了许多全新的统计方法, 如分组法、比较法、推算法、图表法等, 成为社会经济统计学的基本方法。因此, 他被公认为统计学的创始人, 其所代表的学派被称为政治算术学派。约翰·格朗特的代表作是《对死亡率公报的自然观察和政治观察》, 该书对伦敦人口的出生率、死亡率、性别比例做了分类计算并编制了死亡率表, 以死亡率表为依据进行了人口的统计推算预测。因此, 他被认为是人口统计学的创始人。

16 世纪, 意大利学者已经开始研究古典概率论, 但直到 19 世纪初, 才由法国数学家、物理学家拉普拉斯在前人成果的基础上总结了古典概率论的研究成果, 初步奠定了数理统计学的理论基础。拉普拉斯把大数定律作为概率论与政治算术的桥梁, 提出应以自然科学的方法研究社会现象, 为数理统计学的产生提供了必要的理论依据。

## 2. 近代统计学的产生与发展

近代统计学的主要贡献是建设和完善统计学的理论体系, 并逐渐形成了数理统计学和社会统计学两大学派。

(1) 19 世纪前半叶, 资本主义制度在欧洲许多国家中已经成熟, 机械唯物论的世界观和自然科学的成果证实了世界存在自然规律, 这为数理统计学的建立创造了有利条件。

比利时统计学家朗伯·凯特勒在其《社会物理学》一书中将概率论引入统计学, 认为概率论是适用于政治及道德科学中以观察与计数为基础的方法。他以此方法对自然现象和社会现象的规律性进行观察, 并认为要促进科学的发展就必须更多地应用数学。他是古典统计学的完成者, 近代统计学的先驱, 也是数理统计学派的奠基人。

(2) 社会统计学派产生于 19 世纪后半叶的德国。该学派的创始人是卡尔·克尼斯, 他认为统计学是一门独立的、具有政治算术内容的社会科学。另一位有影响的创始人是乔治·梅尔。他把统计学作为实质性研究的社会科学, 并认为统计学以社会集团的规律性作为其独立的研究对象, 以大量观察法作为其特殊的研究方法, 初



拓展窗



步建立了社会统计的学科体系。

社会统计学派的主要代表人物是德国统计学家和经济学家恩斯特·恩格尔，他通过工人家庭生活费用调查发现了“恩格尔定律”。

### 3. 现代统计学的产生与发展

自 19 世纪末以来，欧洲自然科学的飞速发展促进了数理统计学的发展。20 世纪初是推断统计学派发展最迅速的时期，这期间有影响的大师和理论很多，如 20 世纪初戈赛特的  $t$  分布理论，20 世纪 20 年代费喧的  $F$  分布理论，20 世纪 30 年代尼曼等人的假设检验及置信区间估计等理论，20 世纪 40 年代瓦尔德等人的统计决策理论、多元分布理论等。20 世纪 20 年代以后，细胞学的发展使统计学迈进了推断统计学的新阶段。到了 20 世纪 50 年代，经过几代大师的努力，推断统计学的基本框架已经建成，并逐渐成为 20 世纪的主流统计学。

现代主流统计学有以下明显趋势：随着现代数学的发展，人们对数学方法更广泛的应用以及统计学与其他新科学新理论的结合，将不断产生新的边缘科学或新的统计分支；借助计算机，大量数理方法得以普遍应用，计算机已成为实证分析的主要工具；统计的作用，从描述向推断、预测及决策方向发展。

#### 任务卡 3

张美是一个直播卖女鞋的货主，每种款式每次备货最大的是 37 码。请思考其中如何运用了统计学。

### 知识卡 3 统计学的研究对象和特点

#### 一、统计学的研究对象

统计学研究什么，也即统计学的研究对象是什么，这是学习统计学首先要解决的问题。

统计学的研究对象是指统计研究所要认识的客体，它决定着统计学的研究



领域以及相应的研究方法。一般来说,统计学的研究对象是如何认识客观事物的数量特征和数量关系的理论与方法。人们要认识客观事物,就必须通过试验或调查来采集有关数据,并加以整理、归纳和分析,以便对客观事物规律性的数量表现做出统计上的解释。例如,统计需要哪一类数据,怎样采集和加工这些数据,怎样从纷繁复杂的数据中得出结论并解释这个结论,没有统计理论和方法的指导是无法完成的。所以统计学是关于如何采集、整理、显示、描述、分析数据和由数据得出结论的一系列概念、原理、原则、方法和技术的学科。

本书阐述统计学中的社会经济统计学所涉及的统计工作是指社会经济统计工作。我国统计学家在1978年的“统计教学科研规划座谈会”上,将社会经济统计学作为经济领域的统计学,把社会经济统计学、数理统计学和自然科学各领域的统计学确定为统计学的分支。社会经济统计学是一门方法论的科学。社会经济统计学不排斥数理统计方法的运用,数理统计方法应该是社会经济统计学的重要内容。本书着重研究社会经济统计学。社会经济统计的研究对象是大量社会经济现象总体的数量方面,即研究社会经济现象总体的数量特征和数量关系。社会经济统计学的研究对象是关于如何采集、整理、显示、描述、分析大量社会经济现象统计数据 and 由数据得出结论的一系列概念、原理、原则、方法和技术。

## 二、社会经济统计的特点

社会经济统计学从其性质来讲是一种对社会经济现象总体数量方面的认识活动,是一门研究方法论的学科。社会经济统计具有如下特点。

### 1. 数量性

探索总体现象的数量规律、数量特征是统计学的任务,也是统计学最突出的特点。归根结底,统计学的目的是概括出现象总体数量方面的特征和规律,包括数量的多少、现象间的数量关系、现象的质量互变的界限等。

任何社会经济现象都具有质和量两个方面。统计学是对现象总体数量方面的研究,属于定量认识的范畴,但这种定量认识要以定性认识为基础。因为只有对现象的性质、特点、运动过程有一定的认识,才有可能进行定量认识。例如,要了解和分析流动资产投资额的数量、构成及变化,首先必须明确流动资产投资的质的规定性,然后才能根据这种认识确定流动资产投资额的计算范围



和方法。

需要注意的是,统计学中所涉及的数量方面,不论是反映个体属性的统计数据,还是反映总体特征的统计指标,都是客观的、有具体内容的,而不是抽象的数字。这是统计学区别于数学的重要特征之一。例如,2019年我国的GDP为990 865亿元,表明2019年我国所有常住单位生产活动的总成果,而不是一个空洞的数字。

### 2. 总体性

统计学是研究如何收集、整理数据并从中分析概括出关于总体的数量特征、数量规律的方法论学科。尽管统计分析的出发点是从收集反映个体特征的统计数据入手,但最终目的是概括出总体的数量特征和规律。也就是说,统计学是对反映个体的统计数据进行大量观察和综合分析,从而得出反映现象总体的数量特征的。例如,人口统计不是要了解和研究个别的人,而是要反映一个国家或一个地区的人口总数、自然构成、经济构成、社会构成、地域构成、生育率和死亡率、人口迁移流动与城镇化等方面的特征和规律。又如,在某地的消费需求调查中,对每个被调查者进行调查的目的是概括出该地消费者总体的消费需求规律,为企业有针对性地开展营销活动提供依据,每个被调查者只是切入点,并不是统计分析的最终对象。

当然,统计学对现象总体的分析研究,是以对个体属性的认识为基础的。例如,人口统计必须从了解每个人的情况开始,然后经过分组、汇总、计算等工作才能过渡到说明总体数量方面的情况。

### 3. 变异性

统计研究同类现象总体的数量特征,它的前提是总体各单位特征表现存在差异,而且这些差异并不是由某种事先给定的特定的原因造成的。

### 4. 具体性

社会经济统计的研究对象是具体事物的数量方面,不是抽象的量,因此,社会经济统计具有具体性的特点。统计研究的量是在具体时间、地点条件下的量,这个量总是和质紧密地联系在一起的,而数学所研究的量是抽象的量,两者有着明显区别,但统计学中往往借鉴数学的方法。





## 5. 社会性

社会经济统计以社会经济现象作为研究对象，而且社会经济统计本身也就是一种社会实践活动，因此具备社会性的特点。



拓展窗

## 三、统计工作过程

统计工作过程是指统计工作的步骤。统计工作的步骤有统计设计、统计调查、统计整理、统计分析和统计数据提供与管理，如表 1-1 所示。

表 1-1 统计工作的步骤

| 步 骤       | 详 细 说 明                                                                                                                                                                                 |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 统计设计      | 统计设计是根据统计研究的需要和现象的性质，对统计工作的各个方面和各个环节进行全盘计划和安排。统计设计的结果表现为各种统计设计方案，如统计指标体系、分类目录、统计报表制度、调查方案、汇总或整理方案等。统计设计贯穿统计工作的全过程                                                                       |
| 统计调查      | 统计调查是根据统计设计方案的要求，采用科学的方法，对所调查的对象有计划地、系统地收集资料的过程。统计调查是统计整理与统计分析的基础环节。统计调查担负着搜集基础资料的任务，所搜集的资料是否准确关系到统计工作的质量                                                                               |
| 统计整理      | 统计整理是根据统计的目的，采用科学的方法，对调查资料进行加工汇总，使之系统化、条理化的过程。统计整理是统计工作的中间环节，是统计分析的前提                                                                                                                   |
| 统计分析      | 统计分析是对经过加工汇总的统计资料进行分析研究，计算各项综合指标，并利用各种分析方法，揭示现象的数量特征和内在联系，阐明现象的发展趋势和规律性，并根据分析研究得出科学的结论的过程。统计分析是统计工作的决定性环节<br>整个统计过程是统计认识提高的过程，是经过统计设计（质）到统计调查和统计整理（量），再到统计分析（质与量结合），从而达到对现象的本质和规律性的认识过程 |
| 统计数据提供与管理 | 统计数据提供与管理是指将统计整理和分析所得到的统计资料系统化，建立统计数据库和信息网络，以各种方式向社会各界提供信息和咨询。这是统计资料开发利用、统计工作实现自身价值、实现统计信息社会化的重要阶段                                                                                      |

## 四、统计的基本职能

统计的基本职能是指统计本身所固有的内在功能。统计具有信息职能、咨询职能和监督职能。



### 1. 信息职能

信息职能是指统计具有信息服务的功能。也就是说,统计通过系统的收集、整理、分析工作得到统计资料,并在此基础上反复提炼筛选,能提供大量有价值的、以数量描述为基本特征的统计信息,为社会服务。

### 2. 咨询职能

咨询职能是指统计具有提供咨询意见和对策建议的服务功能。也就是说,统计部门要利用所掌握的大量统计信息资源,经过进一步的分析、综合和判断,为宏观、微观决策,为科学管理提供咨询意见和对策建议。同样,统计咨询职能的体现,必须以统计信息的产业化、社会化和国际化为前提条件。统计咨询职能是统计信息职能的延续和深化。

### 3. 监督职能

监督职能是指统计具有揭示社会经济运行中的偏差,促使社会经济运行不偏离正常轨道的功能。也就是说,统计部门要以定量检查、经济监督、预警指标体系等手段,揭示社会经济决策和执行中的偏差,使社会经济决策及执行按客观规律的要求进行。

统计的信息、咨询、监督三大职能是相互作用、相辅相成的,共同构成了统计的整体职能。其中,信息职能是最基本的职能,是咨询和监督职能能够发挥作用的保证;咨询、监督职能是信息职能的延续,两者的强化又会反过来促进信息职能的优化。

发挥统计的整体职能是我国长期统计工作,特别是改革开放以来统计实践经验的总结,是国家科学管理和宏观调控的客观需要。

## 五、统计研究的具体方法

每门学科都有其特定的研究对象,对于不同的研究对象需要用不同的方法去研究。以下是实际生活中经常用到的主要的统计研究方法。

### 1. 大量观察法

大量观察法是统计学所特有的方法。统计学立足于大数定律,认为社会经济现象具有统计规律性,即对现象进行局部观察,其统计数据是随机的、不确定的;但随着观察次数的增加,现象的规律性越来越明显,也就是说,对现象进行大量观察能够挖掘出现象的



拓展窗



规律性。统计史上著名的“抛硬币实验”就反映了这种规律。

大量观察法是指在研究现象和过程时要从总体上加以考察,对现象总体中的全部或足够多的个体进行调查,将充分占有的实际数据资料作为认识的基础。统计学的研究对象是客观现象总体的数量方面,由于社会经济现象总量是复杂的,是在诸多因素的作用下形成的,各单位的特征及数量表现都有很大的差异,不能任意抽取个别或少数单位进行观察,必须观察全部或足够的调查单位,借以从中认识客观现象的规律性。统计调查中的许多方法,如统计报表、普查、抽样调查等,都是通过观察研究对象的大量单位来了解客观事物及其发展规律的。因此,统计学研究的各种现象和过程是就总体中的全部或足够多数量的个体进行调查并综合分析,而不能只取个别或少数个体。

## 2. 统计分组法

根据现象总体的特点及统计研究的目的,把总体按一定的标志划分为不同的组成部分(不同的组),把不同性质的单位分开,而各组内部各单位的性质基本相同,这就是统计分组法。例如,要考查民营企业的亏损面及亏损额,可以选择以“盈亏状况”为标准进行分组,将民营企业分成许多盈亏状况不同的组,从而找出那些亏损严重的企业并分析其存在的问题。统计分组法在统计研究中占有重要地位,它不仅是统计资料整理的重要组成部分,而且贯穿统计工作的全过程。应用统计分组,可以揭示现象的不同类型。在分组的基础上,可以研究现象总体内部构成及内部数量关系。

## 3. 综合指标法

将大量观察所得的资料进行加工、汇总,就可以得到反映现象总体一般数量特征的综合指标。运用各种综合指标对现象总体的数量方面进行分析,这种分析方法称为综合指标法。常用的综合指标有三类:总量指标、相对指标和平均指标。在这三类指标的基础上展开的统计分析的具体形式有对比分析、平均分析、动态分析、变异分析、指数分析和经济模型分析(包括相关回归分析、平衡分析和预测分析)。

## 4. 统计推断法

统计推断法是指以一定的置信水平,根据样本数据资料来判断总体数量特征的归纳推理方法。通常能够进行观察的只有部分或有限单位,而需要判断的





总体对象范围是大量的，甚至是无限的，这样就产生了根据局部的样本数据资料对全部总体数量特征所做判断的置信度问题。例如，要对一批商品的质量进行破坏性检验，只能根据部分商品的质量结果来推断该批商品的质量；根据某市部分职工家庭，如 1 000 户职工家庭的平均收入，可推断该市全部职工家庭的平均收入水平。解决这些问题的方法就是统计推断法。

### 5. 抽样推断法

抽样推断法是在抽样调查的基础上利用样本的实际资料计算出样本数据，并运用概率估计方法推算总体相应的数量指标的一种统计分析方法。对于有些不可能或不必要进行全面调查但又需要了解其全面数量情况的社会经济现象，就可以运用抽样推断法。

### 6. 图表法

图表法是将统计调查得到的凌乱的数字资料加工整理成统计图、统计表，以统计图或统计表的形式反映客观现象的规律性或发展趋势的方法。图表法被广泛应用于统计分析中。

统计的方法仍处于不断的发展中，只要把握住统计研究对象的特点，根据统计发展的需要，吸收一切相关学科有益的研究成果，不断总结统计实践经验，不断完善和发掘新的统计方法，就可以最大限度地发挥统计的信息职能、咨询职能和监督职能。

## 六、统计的组织和管理

随着我国改革开放、建立社会主义市场经济的需要，以及对统计职能的重新认识，统计面临许多新问题。因此，我们应该以《中华人民共和国统计法》（以下简称《统计法》）为基本依据，积极吸取世界其他国家在利用统计管理国家、管理经济，为企业服务、为社会大众服务等方面的先进经验，结合我国的实际情况，积极探索统计改革和发展的新途径。

### （一）统计的组织

统计工作必须贯彻集中统一的原则。我国集中统一的统计系统由各级政府部门的综合统计系统、各级业务部门的专门统计系统，以及城乡基层企业、事业单位的统计组织所组成。另外，统计是一种具有严密的组织纪律，同时由广大统计人员和人民群众参加的工作。如果没有统一的制度和办法，没有统一的



组织领导，就不能保证统计信息的真实性和准确性。

凡属宏观调控所需要的、综合性的、有关基本国情国力的统计信息，都由政府统计部门组织调查，或由政府统计部门与有关业务部门共同组织调查。除此之外，皆由各有关部门自行组织调查，但必须接受政府统计部门的统一管理，不得与国家统计调查重复或矛盾。在现行的部门统计调查项目中，有些可以委托行业协会或统计信息咨询机构承担。基层单位的统计组织，如乡镇统计站和企事业单位统计机构，政府统计系统应进一步加强对它们的组织领导和业务指导，使其在发展农村和城市社会化综合服务方面发挥更大的作用。企业统计信息机构在完成国家和地方统计调查任务的同时，应把工作的重点转移到为企业的生产经营管理服务上来，使企业统计工作保持旺盛的生机和活力。

## （二）统计的管理

统计制度在不同的国家和地区，或者一个国家的不同历史发展阶段，都是不同的。例如，美国没有专门的国家统计局，其统计资料分别由不同的政府部门负责搜集和公布，如有关经济方面的资料由美国商业部提供，劳动工资方面的资料由美国劳工部负责搜集和公布等，这种统计制度和我国是有区别的。下面简要介绍我国的统计管理体制。

### 1. 统计管理体制

中华人民共和国成立后，在中央财经委员会计划局设立了统计处，负责规划全国的财经统计工作。1952年，成立国家统计局，领导全国的统计工作。从那时起，我国建立了由国家统计局，各省、自治区、直辖市统计局，市、县、区统计局等组成的政府专门统计系统和各级政府职能部门及大中型企业内设的专业统计机构，共同进行国民经济和社会各方面的统计工作，实行统一领导、分级管理的管理体制。

我国统计管理体制总的指导思想是：在综合统计系统、专业统计系统和基层单位统计组织组成的总系统下，设立四个子系统，并以此调整各系统的机构设置和职责分工。在总系统内，要注意调动和发挥各子系统的积极性和主动性，处理好集中统一与因地制宜的关系，处理好综合统计与专业统计的关系。政府统计系统的管理体制也可由现行的双重领导体制改为中央统计系统和地方统计系统各自实行垂直管理的体制。在保证新国民经济核算体系正常顺利实施的前



前提下,应该广泛采用计算机等先进的技术手段,最大限度地实现统计信息资源共享,实现统计信息社会化,充分、有效地发挥统计的整体功能。

### 2. 统计调查制度

我国的统计调查制度是从1950年开始的。1950年3月,在中财委统计总处的领导下,我国进行了第一次全国工业普查,此后全国各种统计调查报表制度先后建立起来,而且多是定期报表,如国家工业统计报表主要反映全国工业经济发展的基本情况,为党和国家了解工业经济现状,制定经济政策,编制和检查工业计划执行情况,进行经济管理和宏观调控提供依据;固定资产投资统计报表主要反映全国固定资产投资完成情况,为各级政府制定政策和进行宏观管理提供依据;劳动统计报表主要是为了及时、准确地搜集、整理有关企业(单位)从业人员人数及劳动报酬等方面的资料,为党中央、国务院和各级政府制定有关政策、进行宏观调控及决策提供依据,为国民经济核算提供可靠依据。

从调查方法制度上来讲,过去实行的全面统计报表制度,已不能满足国家经济发展对统计的要求,必须进行改革。以定期普查和抽样调查为主、多种调查方法相结合是我国现行的统计调查体系的基本特征。

### 3. 统计法制

统计法制是保证统计工作依法进行,保证统计资料及时、客观、真实地反映社会经济现实的必要手段。同时,统计法制对于发展市场经济,巩固统计改革成果,保障统计人员的权利,促进社会主义现代化建设事业具有重要的意义。《统计法》于1984年1月1日颁布施行,1987年2月发布《中华人民共和国统计法实施细则》,1996年5月15日发布《关于修改〈中华人民共和国统计法〉的决定》,对《统计法》的基本内容做了具体规定,它为我国在发展社会主义市场经济条件下统计职能的发挥提供了法律上的保证。

认真贯彻、执行《统计法》,就是要依靠法制手段维护国家利益,保护统计调查者、被调查者和信息使用者的合法权益,为引导和推动统计改革和统计现代化建设做出贡献。我国全体公民应自觉执行《统计法》,对违反统计法规的任何组织和个人都必须依法严肃查处,以真正体现法律的严肃性和权威性。



《中华人民共和国统计法》



## 任务二 掌握统计学中常用的基本概念

### 任务卡 1

你了解统计吗？说说你的身边有哪些统计，都用到哪些统计方法。

### 知识卡 1 统计总体与总体单位

#### 一、统计总体

##### 1. 统计总体概述

所谓统计总体，是指在同一性质基础上结合起来的许多个别事物的整体。例如，当研究商业经营情况时，全部商业企业是一个总体，因为商业企业是客观存在的，每个商业企业的经济职能都是相同的，具有同质性，即都是从事商业经营活动的基层单位。

##### 2. 统计总体的特征

在统计调查中，对于有限总体既可进行全面调查，又可进行非全面调查；对于无限总体不能进行全面调查，只能进行非全面调查，根据样本数据推断总体特征。此外，统计总体还可以分为静态总体和动态总体，前者所包含的各个单位属于同一时间，后者所包含的各个单位则属于不同时间。根据一定的目的，针对这两类总体可以分别进行静态研究或动态分析。

总体与总体范围的确定，取决于统计研究的目的和要求。形成统计总体必须具备四个特性：大量性、同质性、差异性和相对性。

(1) 大量性。大量性是大数定律的要求。统计学特有的研究方法是大量观



察法,根据大数定律的思想,要想探寻出总体的数量规律和数量特征,组成总体的总体单位的数量应该足够多,仅仅由个别单位或少量单位不足以显示出总体的规律性。因为个别单位的数量表现可以是各种各样的,只对少量单位进行观察,其结果难以反映现象总体的一般特征。统计研究的大量观察法表明,只有观察足够多的量,在对大量现象的综合汇总过程中,才能消除偶然因素,使大量社会经济现象的总体呈现出相对稳定的规律和特征,这就要求统计总体必须包含足够多的单位。当然,大量性是一个相对的概念,它与统计研究目的、客观现象的规模以及总体各单位之间的差异程度等都有关系。

(2) 同质性。总体的同质性是指构成总体的各个单位具有某种共同性质。同质性是将总体各单位结合起来构成总体的基础。例如,河南工业企业作为统计总体,则每个总体单位都必须具有从事工业生产活动的企业特征,而不具备这些特征的就不能称为工业企业。如果违反同质性,把不同性质的单位结合在一起,对这样的总体进行统计研究,不仅没有实际意义,甚至会产生虚假和歪曲的分析结论。

同质性的概念也是相对的,它是根据一定的研究目的确定的,目的不同,同质性的意义也就不同。例如,研究河南工业企业的生产状况时,所有工业企业都是同质的;而研究民营工业企业的生产状况时,民营工业企业与国有工业企业就是异质的。可见,同质性是相对研究目的而言的,当研究目的确定后,同质性的界限就确定了。

(3) 差异性。差异性是指构成总体的个体除了至少在某一方面具有共同性质外,在其他方面还存在差异。例如,全国的民营企业总体,除了都是中国的民营企业这一共同特点外,在注册资本、投资规模、净利润、职工人数等方面还存在差异。

(4) 相对性。统计总体和总体单位的角色不是一成不变的,两者随着研究目的和任务的不同而不同。同一个客观事物在某项研究中属于个体,但在另一项研究中可能就成为统计总体。例如,在安徽省中职学校这个统计总体中,每个中职学校都是总体单位,安徽新华学校就是其中的一个总体单位。但要研究一个典型中职学校内部的教学科研情况,如果选中了安徽新华学校,那么它就成为统计总体,学校的各院系或每个教职工就是总体单位。





## 二、总体单位

所谓总体单位,是指构成统计总体的个别事物,如前述的每个中职学校就是总体单位。再如,要对全国的人口进行普查,全国人民就是统计总体,而每一个人则是总体单位。根据统计研究目的的不同,总体单位可以是一个地区、一个部门、一个单位,也可以是一个人或一个物。

总体所包含的总体单位数称为总体容量或总体规模,记为 $N$ 。总体按其单位数的多少分为有限总体和无限总体。总体包含的总体单位为有限个,称为有限总体;总体中的单位是无限的或无法计数的,称为无限总体。例如,人口总数、企业总数、商店总数、土地数、各种设备数等都是有限总体;在连续大量生产的某种小件产品中,总产量是无限总体。

### 任务卡 2

卫生间台面与身高有关,因为单位里男子的身高以 172 cm 为最多,人数占 85%,所以台面高度设计就要以他们的身高为参考。请思考卫生间台面的统计标志是什么。

## 知识卡 2 统计标志与统计指标以及统计指标体系

### 一、统计标志

#### 1. 统计标志的含义

标志是说明总体单位属性或特征的名称。例如,全国人口总体,每个人是总体单位,每个人都有性别、年龄、民族、身高、体重等特征。这里的性别、年龄、民族、身高、体重等在统计上就称为标志。这些属性或特征是统计总体各单位身上所共同具有的,都是说明总体单位属性或特征的标志。





## 2. 统计标志的分类

按照不同的划分标准,统计标志可以分为不同的类型。

(1) 按照性质不同,统计标志可以分为品质标志和数量标志。

①品质标志。品质标志是表明总体单位属性的特征,如前面列举的每个同学的性别、民族,其所具有的属性只能用文字说明。例如,性别只能用“男”“女”表示,不能用数值表示,所以性别就是品质标志。

②数量标志。数量标志是表明事物量的特征,用数值来表示的标志。例如,年龄可以用数值(如16岁)来表示,同样,身高、体重等都可以用数值来表示,所以它们均为数量标志。

(2) 按照是否变动,统计标志可以分为不变标志和可变标志。

①不变标志。凡是总体各单位某种标志的具体表现都相同的,这种标志就称为不变标志。例如,要研究某学校教师的工资收入情况,则该学校所有教师便构成总体,“职业”便是其不变标志,是形成总体的前提,即总体的同质性。

②可变标志。凡是总体各单位某种标志的具体表现不相同或不完全相同的,这种标志就称为可变标志。在“职业”为教师的总体中,每位教师的工资收入是不完全相同的,“工资”便是可变标志,它因每位教师的工龄、职称及工作绩效等不同而异。但如果我们研究的不是该校教师的工资收入情况,而是该校所有员工的工资收入情况,总体就应包括教师、行政管理人员、服务人员等在内,这时“职业”这个标志在总体各单位上的表现就不尽相同了,它与工资一样均为可变标志。

## 二、统计指标

### 1. 统计指标的含义

统计指标是反映统计总体数量特征的概念和具体数值。与统计标志不同,它是依附于统计总体的。例如,人口数量、土地面积、工农业生产产量、工业生产总产值、成本、利润、国民收入等概念用于反映一定统计总体的数量方面时就是统计指标。任何统计指标都要通过一定的数值来反映,这种数值就称为指标数值。例如,截至2021年12月,中国60岁以上老年网民规模达1.19亿,互联网普及率达43.2%。

一个完整的统计指标是由两部分构成的,即指标名称和指标数值。指标名



称是反映统计总体现象的概念,表明现象总体的质的规定性,反映某一社会现象内容所属的范围;指标数值则是统计所研究现象的具体数量综合的结果,对某一现象总体特征从数量上加以说明。统计指标名称及其指标数值的有机结合,也就是事物质的规定性和量的规定性有机联系的表现。

## 2. 统计指标的分类

按照不同的划分标准,统计指标可以分为不同的类型。

(1) 按说明的总体现象的内容不同,统计指标可以分为数量指标和质量指标两类。

①数量指标是说明现象总体绝对数量多少的指标,它反映的是总体外延的广度、规模大小及其发展成果多少的总和。它用绝对数来表示,并有实物或货币的计量单位。例如,耕地面积、人口数、国内生产总值等都是数量指标。数量指标受总体范围的影响,它的数值随总体范围的大小而增减,是认识总体现象的出发点。

②质量指标是说明总体内部数量关系和总体单位水平的指标,反映现象本身质量、现象的强度、经营管理工作质量和经济效果等的统计指标,用来说明总体的质的属性。质量指标表示事物的内涵量状况,如产品合格率、劳动生产率、人的性别和年龄构成等都属于质量指标。质量指标用相对数或平均数表示。由于质量指标反映的是现象总体内部的数量关系,因此其指标的数值不随总体范围的大小而增减。

数量指标是计算质量指标的基础,质量指标往往是相应的数量指标进行对比的结果。在实际工作中,要把数量指标和质量指标结合起来应用。在研究总体现象时,不仅要用数量指标了解总体的绝对数量多少及其发展变化,而且要用质量指标了解总体内部的数量关系及其发展变化,两者不可偏废。

(2) 按作用和表现形式不同,统计指标可以分为总量指标、相对指标和平均指标。

①总量指标是反映总体现象规模的统计指标,是说明总体现象广度的,它表明总体现象发展的结果。总量指标的数值随总体范围的大小而增减,并具有可加性。总量指标可以反映一个国家和地区国情国力的基本情况,是制定政策、编制计划的基本依据,如人口总数、土地面积、国民生产总值等。



②相对指标是两个有联系的总量指标相比较的结果,反映总体之间或总体内部各组成部分之间的数量关系,如产品产量的计划完成程度、人口密度、人口的年龄构成等。相对指标在数值上与总体范围的大小无直接的相关关系,不具有可加性。

③平均指标是按某个数量标志表明同类社会经济现象在一定时间、地点条件下所达到的一般水平,如平均工资、人均土地面积、平均计划完成程度等。同相对指标一样,平均指标在数值上与总体范围的大小无直接的相关关系,不具有可加性。

由于事物的内容决定其形式,因此,我们所说的总量指标是数量指标的表现形式,相对指标和平均指标是质量指标的表现形式。

(3)按在管理上所起的作用不同,统计指标可以分为考核指标和非考核指标。

①考核指标是根据管理的需要,用来考核成绩、评定优劣、决定奖罚的统计指标。它是从所有统计指标中精选出来的若干统计指标。它的现实作用很大,直接影响地区、部门、单位以及劳动者的荣誉、物质利益和积极性。因此,这类指标的确定必须严肃、认真、细致,对其含义、界限、计算方法等都要规定得明确而又具体,一般来讲不宜过多。

②非考核指标是用于了解情况和研究问题的统计指标。在一个单位,一般来说,非考核指标的数量要多于考核指标的数量,因而也不能忽视。



拓展窗

### 三、统计指标体系

#### 1. 统计指标体系的含义

统计指标体系是由若干个相互联系、相互作用的统计指标组成的整体,用以说明所研究社会经济现象各方面相互依存和相互制约的关系。例如,一个工业企业把产品产量、净产值、劳动生产率、质量、消耗、成本、销售收入等统计指标联系起来就组成指标体系,这便于全面、准确地评价该企业的生产经营情况。

#### 2. 统计指标体系的形式

统计指标体系通常表现为以下两种情况。



(1) 可以通过数学公式形式表现出来的统计指标体系。

总产值 = 产量 × 单价

销售额 = 销售量 × 单价

总成本 = 产量 × 单位成本

原材料总消耗额 = 产量 × 单耗 × 单价

(2) 指标之间不存在数学公式形式的关系，而只存在一种相互联系、相互补充的关系。例如，反映国内商品流转情况的指标（购进量、销售量、调拨量、库存量）所形成的指标体系，考核商业企业经济效益的指标（劳动生产率、人均利税率、资金利税率、费用利税率、流动资金周转次数）所组成的指标体系都属于这种情况。

### 3. 统计指标体系的分类

(1) 按照统计指标体系的反映内容分类。统计指标体系按其反映内容的不同，可分为社会统计指标体系、经济统计指标体系和科学技术统计指标体系。它们分别从人口社会、国民经济运行和科学发展三个方面，反映一个时期、一定范围内国民经济和社会科技发展的总体状况。

(2) 按照统计指标的考核范围分类。统计指标体系按其考核范围不同，可分为宏观指标体系、中观指标体系和微观指标体系。宏观指标体系反映整个社会、经济和科技情况；中观指标体系反映各个地区和各个部门、行业的社会、经济和科技情况；微观指标体系反映各企业、事业单位的生产经营或工作运行情况。

(3) 按照统计指标的作用功能分类。统计指标体系按其作用功能不同，可分为描述性指标体系、评价性指标体系和预警性指标体系。描述性指标体系主要反映社会经济现象的现状、运行过程和结果；评价性指标体系主要比较、判断社会经济现象的运行过程、结果是否正常；预警性指标体系对经济运行过程进行监测，起预警作用。



拓展窗



### 任务卡 3

有一个袋子，里面有 8 个黑球，12 个白球，它们除了颜色不同外，其他方面没有区别。现将球随机地一颗颗摸出来，则第 10 次摸出的球是黑球的概率是多少？其中的变量是什么？

---

---

## 知识卡 3 变异与变量

### 一、变异

#### 1. 变异的含义

标志在各单位的具体表现互有差别，还在不同时空上有数值的差异，统计学上把这种差别与差异称为变异。统计中的标志有可变标志和不变标志，可变标志就是标志的具体表现各不相同，如性别标志表现为男、女，年龄标志表现为 25 岁、26 岁等，这种差别就是变异。变异包括质的差别和量的差别。变异是统计的前提条件，没有变异就无所谓统计。

#### 2. 变异的分类

变异分为品质变异和数量变异。品质标志在总体各单位之间的具体表现不同称为属性变异，如企业有国有企业、合资企业、私营企业等组织形式上的差别。数量标志在总体各单位表现的差异称为数量变异，如不同企业在利润上有多少的差异。

### 二、变量

#### 1. 变量的含义

变量是指可变的数量标志和指标。变量的具体表现称为变量值，也称标志值。如上网者的年龄分别为 17 岁、18 岁或 60 岁等；各企业的利润分别为 40 万元、350 万元、22 亿元等。



## 2. 变量的分类

(1) 按照数据的连续性不同,变量可以分为连续型变量和离散型变量。连续型变量是指统计在描述其对象基本特征时,某一变量在一定区间范围内可取无限多个数值的变量,如身高、企业利润等。离散型变量是指统计在描述其对象基本特征时,某一变量在一定的区间范围内可以按一定的顺序一一列出变量值的变量,如职工人数、机器台数等。

(2) 按照在研究中的作用不同,变量可以分为自变量、因变量和外变量。自变量是在研究中作为假设的引起某种现象的“原因”的变量。因变量是在研究中假设的“后果”,是受另一个变量(自变量)影响和诱发而产生的。外变量是指研究中可能伴随主要变量出现的研究对象的其他特性,可能会对因变量产生积极或消极的影响,因此是研究中应尽量控制的因素。



统计学发展历程

## 职业基础知识测试

### 一、填空题

1. 社会经济统计的特点分别是\_\_\_\_、\_\_\_\_、\_\_\_\_、\_\_\_\_、和\_\_\_\_。
2. 统计工作的步骤分别是\_\_\_\_、\_\_\_\_、\_\_\_\_、\_\_\_\_、\_\_\_\_、\_\_\_\_。
3. 若要了解某食品企业的产品销售情况,总体是\_\_\_\_,总体单位是\_\_\_\_。
4. 统计的职能分别是\_\_\_\_、\_\_\_\_、\_\_\_\_、\_\_\_\_。

### 二、简答题

1. 社会经济统计的研究对象是什么?
2. 统计研究的具体方法有哪几种?
3. 说说你的身边有哪些统计,都用到哪些统计方法。
4. 什么是标志? 什么是指标?





### 【实训目标】

加强理论和实践的联系，提高学生将统计方法应用于日常学习与生活的能力。

### 【具体要求】

1. 全班 42 名同学本门课程的期中考试成绩（单位：分）如下：

88, 87, 88, 58, 66, 87, 77, 78, 74, 73, 69, 64, 79, 75, 98, 93, 84, 83, 85, 83, 55, 53, 30, 49, 71, 72, 83, 82, 84, 85, 92, 90, 84, 88, 86, 76, 76, 75, 88, 89, 89, 90。

这些成绩是全班每个同学的个体数据，仅是原始资料的堆积。从这些数据我们看不出全班同学这个总体的成绩的基本状况，如全班总平均分是多少，及格人数是多少，优秀人数是多少，以及各分数段的人数占比是多少。

2. 学校规定：60 分以下为不及格，60~70 分为及格，70~80 分为中，80~90 分为良，90~100 为优。

### 【实训记录】

### 【交流与评价】

| 项 目  | 成 绩             |                 |                 |     |
|------|-----------------|-----------------|-----------------|-----|
|      | 统计速度<br>( 30% ) | 课堂氛围<br>( 30% ) | 正确结果<br>( 40% ) | 总 分 |
| 自我评价 |                 |                 |                 |     |
| 同学评价 |                 |                 |                 |     |
| 教师评价 |                 |                 |                 |     |



## 项目二 了解统计调查

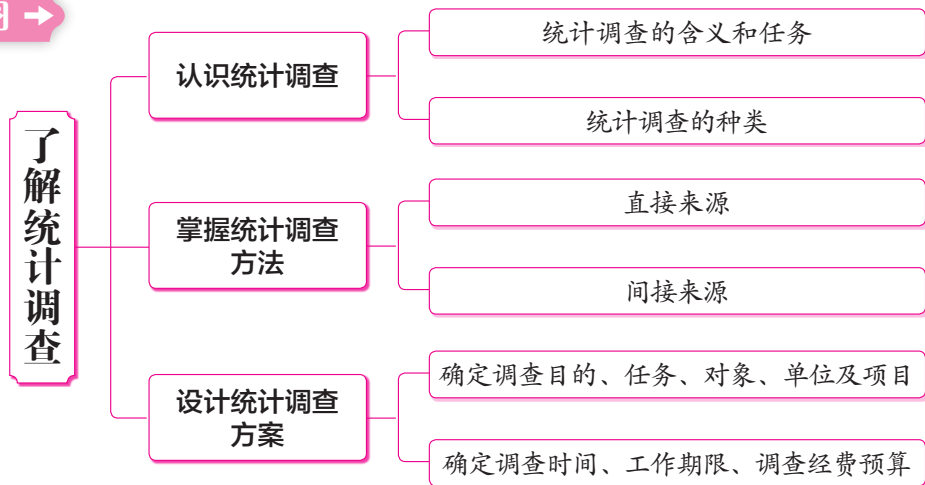
### 知识目标 →

- 了解统计调查的含义、任务及种类
- 理解我国常见的几种统计调查方式，掌握普查、典型调查、重点调查和抽样调查的概念和特点
- 理解统计调查方案的内容

### 能力目标 →

- 学会运用统计调查的具体方法
- 能够利用统计方法完成统计设计工作

### 思维导图 →



### 1. 目的

为更好地服务业主，广泛听取业主意见，及时了解业主对物业服务的满意程度，改进服务内容，提高服务质量，特制定本调查方案。

### 2. 职责

(1) 项目部经理负责服务区业主满意度调查的组织工作。

(2) 项目部经理助理负责具体满意度调查工作的落实，汇总工作。

(3) 物管员负责责任区满意度调查工作的实施工作。

### 3. 工作程序

(1) 标准和要求。

①项目部每年至少对业主进行一次业主意见调查（可根据项目部需要增加次数）。新办理入住的项目部需在办理入住半年后对业主进行一次调查。

②业主满意度调查表的发放数量应达到项目已办理入住户数的90%，回收数量应达到发放数量的80%，整体满意度应达90%（含）以上，各分项的满意度应达到90%（含）以上。

③项目部负责将最终的调查结果进行汇总及各项分析，并报公司综合管理部备案。

(2) 项目部经理助理负责制定本服务区《业主满意度调查表》，负责对最终业主满意度结果的汇总工作。

(3) 物管员根据项目管理情况，向已办理入住的业主发放《业主满意度调查表》，进行满意度调查，调查的内容涉及各部门员工的礼仪礼貌、服务态度、服务质量、处理问题的效率及业主对物业公司的其他意见等。

(4) 调查表回收后，项目部经理助理依据调查结果进行分析，并依据发放量及回收量统计满意度，填写《业主满意度调查结果》。

①业主满意度调查分四个等级：非常满意、满意、一般、不满意。

②发放数量计算方法：已办理入住数量  $\times 90\%$  = 发放数量。

③回收数量计算方法：发放数量  $\times 80\%$  = 回收数量。

④整体满意度计算方法：满意数量（包括非常满意和满意） $\div$  回收数量  $\times 100\%$  = 整体满意度。

⑤分项满意度计算方法：分项满意数量  $\div$  回收数量  $\times 100\%$  = 分项满意度。

（5）各项目部依据问题汇总，制定进一步的整改措施和预计完成时间，报公司总经理审核。

（6）项目部将经公司总经理审批后的整改措施及预计时间反馈业主。

（7）业主满意度调查结束后，由项目部经理助理对此次调查中不满意业主进行回访，达到业主满意。

（8）如业主在调查中提出投诉事宜，应执行“业主投诉处理程序”。

（9）在项目部对业主提出的问题进行整改后，由物管员针对整改结果对业主进行回访并填写“业主满意度调查回访记录”。

（10）满意度调查及对调查的回访工作完成后，调查表及相关记录由项目部统一存档。

（11）注意事项。

①项目部负责人对业主满意度调查的全过程起到监督作用，保证满意度调查的真实性。

②业主满意度调查表发出后，应及时进行回收。

③对业主在意见调查提出的问题，各部门应对改进情况进行跟进，并做详细记录，同时应将最终完成时间在《业主意见调查问题及整改措施汇总》中记录明确。

此调查方案的设计规范吗？有哪些优点和缺点？请上网查找资料，说说你的看法。

---

---



## 任务一 认识统计调查

### 任务卡 1

最近某学校新开设了一个食堂，里面有各种各样的食物，但是就餐的学生并不多。为了查找原因，校长决定让助理以问卷的方式调查学生对食物的满意程度。假如你是校长助理，你会如何设计调查问卷？

### 知识卡 1 统计调查的含义和任务

#### 一、统计调查的含义

统计调查是根据统计任务的要求，运用科学的调查方法，有计划、有组织地向社会搜集统计资料的过程。统计调查是统计工作的基础环节，是认识事物的起点。

通过统计调查收集的资料有两种：一种是对调查资料未做任何加工整理的原始资料，又称为初级资料；另一种是次级资料，即已经经过某个部门或地区加工整理的说明某个部门或地区综合情况的统计资料。

#### 二、统计调查的任务

统计调查的基本任务是根据统计指标和指标体系，通过具体的统计调查，取得反映社会经济现象及其内在联系的原始统计资料。

统计调查担负着为整个统计工作提供基础资料的任务。在统计调查中，对统计资料的收集必须做到准确、及时、全面、系统，否则就不能很好地发挥统计认识社会的作用，甚至会导致错误的结论，造成严重的后果。



### 任务卡 2

要判断一锅汤的味道不需要把整锅汤都喝完，将锅里的汤搅拌均匀，品尝一小勺就知道汤的味道，这是一个简单随机抽样问题。

你同意上面的观点吗？说说你的看法。

## 知识卡 2 统计调查的种类

### 一、按调查对象的范围分类

按调查对象的范围不同，统计调查可分为全面调查和非全面调查。

#### 1. 全面调查

全面调查是对被调查对象中所有的单位全部进行调查，其主要目的是取得总体的全面、系统、完整的总量资料，如普查和统计报表。全面调查要耗费大量的人力、物力、财力和时间。



想一想

#### 2. 非全面调查

非全面调查是对被调查对象中一部分单位进行调查，如重点调查、典型调查、抽样调查和非全面统计报表等。

全面调查和非全面调查是以调查对象所包括的单位范围的不同来区分的，而不是以最后取得的结果是否反映总体特征的全面资料来区分。



拓展窗

### 二、按调查时间是否连续分类

按调查时间是否连续，统计调查可分为经常性调查与一次性调查。

#### 1. 经常性调查

经常性调查是随着调查对象在时间上的发展变化而随时对变化的情况进行连续不断的记录，其主要目的是获得事物全部发展过程及其结果的统计资料。





## 2. 一次性调查

一次性调查是不连续登记的调查，它是对事物每隔一段时期后在一定时点上的状态进行登记，其主要目的是获得事物在某一时点上的水平、状况的资料。

一次性调查又分为定期和不定期两种。定期调查是每隔一段固定时期进行一次调查。不定期调查是时间间隔不完全相等。

## 三、按调查的组织方式分类

按调查的组织方式不同，统计调查可分为统计报表和专门调查，其中，专门调查又可分为普查、抽样调查、重点调查和典型调查等。

### 1. 统计报表

统计报表是指按照国家有关法律法规的规定，自上而下地统一布置、自下而上地逐级提供基本统计数据的一种调查方式。

统计报表是收集统计数据的一种重要方式。在我国几十年的政府统计工作中，已经形成了一套比较完备的统计报表制度，它成为国家和各级地方政府部门统计数据的重要来源。

统计报表有多种分类方法：按报送范围不同可分为全面统计报表和非全面统计报表；按报送周期不同可分为日报、周报、旬报、月报、季报、半年报、年报等；按报送的单位不同可分为基层报表和综合报表；按主管机关、报表内容和实施范围的不同，可分为国家统计报表、部门统计报表和地方统计报表。

### 2. 普查

普查是一种专门组织的一次性的全面调查，用来调查属于一定时点或时期内的社会经济现象的总量。普查要遵循以下原则：

（1）应确定普查的统一时间。

（2）普查的登记工作应在整个普查范围内同时进行，以保证普查资料的实效性、准确性，避免资料的收集工作拖得太久。

（3）同类普查的内容和时间在历次普查中应尽可能保持连贯性。普查的组织形式有两种：一种是组织专门的普查机构，派专门的调查人员对被调查单位直接进行登记；另一种是不设立统一的组织机构，也不配备专门的普查人员，而是利用企业、机关、事业单位本身的组织系统和内部的原始记录、报表资料进行填报，或者结合清库盘点进行登记和调查。



### 3. 抽样调查

抽样调查是从总体中抽取一部分单位作为样本进行观察,并根据观察的结果推断总体数量特征的一种非全面调查方法。

随机抽样是一种重要的抽样方法。随机抽样一般是指每个总体单位都有同等被抽中的机会,但是在实际调查中,并不完全是这种情况。通常采用的抽样组织形式如表 2-1 所示。

表 2-1 通常采用的抽样组织形式

| 组 织 形 式 | 详 细 说 明                                                                                                                                |
|---------|----------------------------------------------------------------------------------------------------------------------------------------|
| 简单随机抽样  | 简单随机抽样又称纯随机抽样,它是指对总体不做任何处理,不进行分类也不进行排除,而是完全按随机的原则,直接从总体中抽取样本单位加以观察。从理论上说,简单随机抽样是最符合抽样调查随机原则的方式,是抽样调查的最基本形式                             |
| 分层抽样    | 分层抽样又称类型抽样或分类抽样,它是指先将总体各单位按主要标志加以分层,然后在各层中按随机的原则抽取若干样本单位,由各层的样本单位组成一个样本                                                                |
| 等距抽样    | 等距抽样又称机械抽样或系统抽样,它是指将总体全部单位按某一标志排队,然后按固定的顺序和相等间隔在总体中抽取若干样本单位,构成一个容量为 $n$ 的样本                                                            |
| 整群抽样    | 整群抽样是指将总体各单位划分为若干群,然后以群为单元,从总体中随机抽取一部分群,对被抽中的群内所有单位进行全面调查。整群抽样对总体划分群的基本要求是:第一,群与群之间不重叠,即总体中的任意单位只能属于某个群;第二,全部总体单位毫无遗漏,即总体中的任意单位必须属于某个群 |
| 多阶段抽样   | 多阶段抽样是指当总体很大时,可把抽样过程分成几个过渡阶段,到最后才落实具体的样本单位                                                                                             |

### 4. 重点调查

重点调查是一种非全面调查,即在调查对象中选择一部分对全局具有决定性作用的重点单位进行调查。重点调查适用于调查任务只要求掌握调查总体的基本情况,调查标志比较单一,调查标志表现在数量上集中于少数单位,而这些少数单位的标志值之和在总体中又占绝对优势的情况。

重点调查的方式有两种:一种是专门组织的一次性调查,另一种是利用定期统计报表经常性地对一些重点单位进行调查。重点调查的优点是花费较少人



力、物力，在较短时间内及时取得有关的基本情况。

## 5. 典型调查

典型调查是根据调查的目的与要求，在对被调查对象进行全面分析的基础上，有意识地选择若干具有典型意义的或有代表性的单位进行调查。典型调查的主要作用：第一，补充全面调查的不足；第二，在一定条件下验证全面调查数据的真实性。

典型调查的优点是灵活机动，通过少数典型即可取得深入、翔实的统计资料，缺点是受“有意识地选出若干有代表性”的限制，易受人们主观认识的影响，必须同其他调查结合起来使用，才能避免出现片面性。



拓展窗

## 任务二 掌握统计调查方法

小明：刚才在统计调查的种类划分时提到的人口普查以及城乡青少年网民上网设备对比的小贴士都是比较直观的例子，只是政府是通过什么方法进行人口普查呢？我爸爸正在和一些专业调查机构合作，这样的专业调查机构又是如何统计城乡青少年网民上网设备对比的呢？和人口普查用的方法一样吗？

### 任务卡 1

某学校的黑板报上刊登了一篇题为《我校大部分学生不吃早餐》的文章。文章说：“本报小记者通过对课间到学校商品部买小食品的 20 名学生的调查，发现有 16 人是因为没有吃早餐而去买零食。由此推断，我校 80% 的学生不吃早餐。”你认为该调查结论正确吗？为什么？

---

---



## 知识卡 1 直接来源

### 一、访问法

访问法就是调查者与被调查者直接或者间接接触，以询问回答的方式取得数据的一种调查方法，具体包括面谈访问、座谈会、电话调查、网络调查等。

#### 1. 面谈访问

面谈访问是调查者与被调查者通过面对面交谈的方式取得数据的调查方法。面谈访问有结构式访问和非结构式访问两种。

结构式访问是指被调查者按照事先设计好的、有固定格式的标准化调查问卷或表格回答问题。结构式访问中的问卷和表格有自填式（由被调查者根据卷面自行填答）和访问式（由调查者提问、被调查者作答的方式收集资料）两种。结构式访问的优点是能够对调查过程加以控制，从而获得比较可靠的调查结果，调查结果便于归类、汇总、分析。

非结构式访问是指事先不制作统一的问卷或表格，没有统一的提问顺序，调查者只是给出一个题目或提纲，由调查者和被调查者自由交谈，以获取所需资料的调查方法。非结构式访问的优点是调查内容的自由度大，如果引导得当，可以获得大量有用信息；缺点是调查结果难以归类汇总，不便于统计分析。

#### 2. 座谈会

座谈会又称焦点访谈法，是指挑选一组有代表性的被调查者集中于调查现场，在主持人的组织下，就某个主题进行讨论、发表意见，从而获得所需资料的调查方法。

#### 3. 电话调查

电话调查是以打电话的方式从被调查者那里取得资料的调查方法。电话调查是随着电话在普通家庭中的普及而迅速发展起来的调查方法。其优点是操作方便，资料回收率高，便于随机抽样，调查的成本低；缺点是对调查者的要求较高，调查内容难以深入全面。

#### 4. 网络调查

网络调查是传统调查在新的信息传播媒体上的应用。它是指在互联网上针



对特定的问题进行的调查设计、收集资料和分析等活动。与传统调查方法相类似，网络调查也有对原始资料的调查和对二手资料的调查两种方式。

根据调查方法的不同，网络调查可分为网上问卷调查、网上讨论、网上实验和网上观察等。

## 二、观察法

观察法是调查者在现场对被调查者的情况直接观察、记录，以取得信息资料的一种调查方法。

观察法可从不同的角度进行分类。

(1) 按照观察者的身份，可以把观察法分为参与观察和非参与观察。参与观察是指观察者直接加入某个群体，以内部成员的角色参与他们的各项活动，在共同生活中观察和收集资料。非参与观察是指观察者以旁观者的身份，置身于调查群体之外进行的观察。

(2) 按照标准化程度，可以将观察法分为结构式观察和非结构式观察。结构式观察是事先制订好观察计划并严格按照规定的内容和程序实施的观察。其优点是标准化程度高，便于操作实施，也便于归类分析。非结构式观察是指对观察的内容、程序事先不做严格规定，依据现场的具体情况随机决定观察。非结构式观察的优点是灵活机动，信息量大，如果观察者的能力强素质高，可探寻出现象内在的属性；缺点是对观察员的素质和能力要求高，所得资料不系统、不规范，难以归类、分析。

(3) 按照观察的方式，可以将观察法分为直接观察和间接观察。直接观察是指观察者直接对被观察者的活动进行观察。间接观察是指通过对自然物品、社会环境、行为痕迹等进行观察，以便间接反映被调查者的状况和特征。例如，通过对某个城市市容卫生的观察，就能从侧面了解该市居民的精神面貌。

## 三、实验法

### 1. 实验法的含义

实验法是根据自然科学中科学实验的原理，从影响调查对象的若干因素中选出一个或几个因素作为实验因素，在其余因素不变的条件下，了解实验因素的变化对调查对象的影响程度。



## 2. 实验法的分类

实验法可以分为以下几类：

(1) 对照实验。一种新药问世时，研究人员即使在实验室进行了大量实验，仍有可能不清楚这种新药对于患者是否有明显的疗效。怎样设计一个实验来测试它的效果呢？最基本的办法就是比较。用对照实验的方法来测试新药的效果，整个实验在双盲情况下进行。双盲是指不管是病人还是测定反应的医生都不应知道谁在处理组，谁在对照组。首先按照随机的方式将病人分为处理组和对照组，处理组由使用新药的病人组成，对照组由不使用新药的病人组成，然后比较两组的反应。为了避免偏差，处理组与对照组除了处理这一因素外应尽可能相似。

(2) 观测研究。观测研究不同于对照实验。观测研究的被观测对象不是由研究人员安排到不同组去的，而是他们自己到不同的组中去，研究人员只是观测所发生的情况。在观测研究中，研究人员要注意对混杂因素的控制。

(3) 销售实验。销售实验就是先在一个较小的范围内，并在一定的条件下（类似实验室的条件下），对某种影响商品销售的因素进行实际实验，分析其结果，以判断这种方法是否有大规模推行的价值，所以这种实验也常称为销售实验，或者称为试销。对试销效果的调查，需要限定在一个特定市场，也就是实验市场。

实验调查法的应用范围很广，凡是某种商品在改变它的质量、包装、设计、价格、广告宣传和陈列方法等因素时都可应用实验调查法，在实验市场上推出做了改变的产品，与未做改变的产品的对照之后就可摸清顾客的反应。

(4) 改革实验。建立社会主义市场经济体制需要经验，经验不足，就需要通过实验获得较为准确的基础数据，如深圳经济特区的建立就采用了此方法。

### 任务卡 2

在一场篮球比赛的实况转播中，解说员介绍了参加美国职业篮球比赛（NBA）的3名中国籍选手的身高。有位观众把这三个人的平均身高与美国球员的平均身高进行比较，得出了一个结论：中国人的平均身高比美国人高。你认为该调查结论正确吗，为什么？

---

---





## 知识卡 2 间接来源

### 1. 查阅公开出版物

在我国，公开出版的社会经济统计数据主要来自国家和地方的统计部门以及各种报刊媒体，如《中国统计年鉴》《中国市场统计年鉴》以及各省、市、自治区的统计年鉴，提供世界各国社会和经济数据的《世界经济年鉴》《世界发展报告》等。

### 2. 向政府统计机构咨询

政府统计机构除了定期公布或公开出版数据资料外，还有相当一部分数据资料未公开或出版。在不涉及国家机密、商业机密和个人家庭隐私的情况下，信息需求者也可咨询了解。

### 3. 向其他机构咨询

信息需求者还可以通过有偿咨询等手段向学术研究机构、商业性统计调查机构收集各种统计调查数据。

### 4. 网络查询

随着计算机网络技术的发展，信息需求者还可以在网络上获取所需的各种数据资料，但是要注意的是，为了保证数据的真实性和权威性，一定要选择具有权威性的网站，如国家和地方统计部门的专业网站，并对数据进行仔细核实。另外，还要注意使用合法的手段获取。



## 任务三 设计统计调查方案

### 任务卡 1

一家上市公司的总裁准备新招聘一个统计管理岗位的工作人员。在面试环节，几位前来应聘的人就设计统计调查方案分别表达了各自的观点。

A：确定统计调查的目的和任务是制订调查方案的首要问题。

B：在调查的目的和任务确定之前，先要确定调查对象和调查单位。

C：调查项目是指调查单位所要调查的主要内容。

D：调查项目确定后，就可以按照调查任务，将调查项目制作成调查问卷或调查表。

假设你是该公司总裁，你比较赞同哪位的观点？为什么？

---

---

### 知识卡 1 确定调查目的、任务、对象、单位及项目

#### 一、确定调查目的和调查任务

确定统计调查目的和任务是制订调查方案的首要问题。对任何社会经济现象的研究，都可以根据不同的目的、任务从不同的角度收集资料。如消费者调查，可以从消费需求的角度进行调查，也可以从消费行为或消费心理等角度进行调查，这就需要明确本次调查的目的是什么。调查的目的和任务不同，调查的内容和范围也就不同。目的不明、任务不清，就无法确定向谁调查，调查什么，以及用什么方式方法进行调查。这就会使调查工作带有很大的盲目性，收集的资料可能是不需



拓展窗



要的,需要了解的情况得不到充分的反映。这样不仅会造成人力、物力、财力的大量浪费,而且会延误工作。

## 二、确定调查对象和调查单位

调查目的和任务确定以后,就要确定调查对象和调查单位。调查对象是需要调查的某一社会经济现象的总体,它由性质上相同的许多调查单位所组成。调查单位是构成社会现象总体的个体,是调查项目的具体承担者,也就是在调查过程中所要调查的具体单位。例如,调查目的是收集某市中职学校状况的资料,那么该市所有中职学校就是调查对象,而全市的每一所中职学校就是调查单位。又如,调查目的是收集私营食品企业的设备资料,则调查对象是一切私营食品企业的所有设备,而私营食品企业的每一台设备则是调查单位。

确定调查对象和调查单位时要注意以下问题:

(1) 由于经济现象具有复杂多变的特点,因此,在许多情况下调查对象是比较复杂的,必须根据研究目的严格规定调查对象的含义,指出它与其他有关现象的区别,以免造成调查中由于界限不清而发生的差错。如“××市城市居民家用计算机消费观念调查”,以××市城市居民为调查对象,就要明确“城市居民”的含义,划清“城市居民”和“非城市居民”概念的界限。

(2) 调查单位的确定取决于调查目的和调查对象,调查目的和对象变了,调查单位也要随之改变。例如,要进行××市城市居民家庭家用计算机消费现状调查,调查单位是××市每一户城市居民家庭,而不再是每一个城市居民了。

(3) 不同的调查方式会产生不同的调查单位。如果采用普查方式,则总体内所有单位都是调查单位;如果采取抽样调查方式,则抽取出的样本单位是调查单位。

## 三、确定调查项目

调查项目是指调查单位所要调查的主要内容。在进行调查之前,必须根据调查目的,明确规定出调查项目。一般来说,确定调查项目应注意以下几点:

(1) 确定调查项目,要有取得资料的可能性。凡是被确定为调查内容的项目,必须能够取得确切的资料。否则,即使需要,但没有条件取得确切资料的项目,也不该列入。同时,对每一个调查项目都应该有确切的含义和统一的解释,以免因为调查人员理解不同而致使调查结果不一致。



(2) 被确定为调查内容的每个项目, 应该彼此衔接, 以便对现象的相互联系从整体上了解, 也便于有关项目相互核对, 提高调查资料的质量。

(3) 调查项目之间时间上要有可比性, 即本次调查项目和过去同类调查项目之间要互相衔接, 以便进行动态对比。

(4) 能确定的项目必须与调查目的有关。只登记与问题有关的标志, 不应包括可有可无、备而不用的标志。

调查项目确定后, 就可以按照调查目的和要求, 将调查项目制作成调查问卷或调查表。

### 任务卡 2

如果你是一家上市公司的统计管理员, 公司下个月要对产品合格率进行调查, 你会如何预估调查经费? 在预估时应考虑哪些方面?

## 知识卡 2 确定调查时间、工作期限、调查经费预算

### 一、确定调查时间和工作期限

调查时间是指调查资料所属的时间。如果所要调查的是时期现象, 就要明确规定该现象是从何年何月何日起到何年何月何日止。如果所调查的是时点现象, 就要明确规定统一的标准调查时点。

为了保证调查工作的及时性和有序性, 还要规定调查工作期限。调查工作期限是指进行调查工作的时限, 包括从方案设计、搜集资料到提交调查报告的整个工作进度。规定调查期限要考虑调查项目的复杂性和调查资料的时效性。在可能的情况下, 调查期限应尽可能缩短。

### 二、确定调查经费预算

企业所进行的市场调查, 要涉及调查经费的筹措和使用, 对费用进行估算

也是调查方案的内容之一。调查费用的多少要视样本容量、调查范围以及调查内容的复杂程度而定，没有绝对的标准。进行费用估算时，一般要考虑调查方案设计费、抽样费、问卷设计费、问卷印刷费、调查费（培训费、交通费、调查人员工资、礼品费等）、数据录入费、数据统计分析费、调查报告费、办公费和其他费用等。

### 三、制订调查组织实施计划

这是调查工作顺利实施的保证。调查的组织工作主要包括以下内容：调查项目组的构建、调查员的组织和培训、调查前的准备工作（包括调查动员、资料查询、文件印刷等）、调查经费的筹措和管理制度的制定等。

调查方案的设计是调查工作的出发点，也是整个调查活动的行动指南，务必站在全局的角度对调查工作的各环节、各阶段进行通盘考虑，制订出合理的工作计划。



## 素质园地

## 职业基础知识测试

### 一、填空题

1. 统计调查的基本任务是根据\_\_\_\_\_，通过具体的\_\_\_\_\_，取得反映社会经济现象及其内在联系的\_\_\_\_\_。
2. 小明在路上走，一位姐姐拦住他并想让他配合做一个调查，这属于调查方法中的\_\_\_\_\_。在超市，小明看到一位老爷爷被邀请坐在按摩椅上体验并回答几个问题，这属于调查方法中的\_\_\_\_\_。
3. 全面调查是对被调查对象中\_\_\_\_\_进行调查，其主要目的是取得总体的全面、系统、完整的\_\_\_\_\_。
4. 普查要遵循以下原则：确定普查的\_\_\_\_\_；普查的\_\_\_\_\_应在整个普查范围内同时进行，以保证普查资料的实效性、准确性，避免资料的收集工作拖延太久；同类普查的内容和时间在历次普查中应尽可能保持\_\_\_\_\_。
5. 统计调查方法分为直接来源和间接来源，直接来源又分为\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_和网络调查，间接来源又分为\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_。



## 二、简答题

1. 统计调查有哪些分类方法？
2. 简述抽样调查的含义。
3. 你做过哪些统计调查？有什么体会与同学分享？
4. 怎样设计统计调查方案？



## 职业核心能力训练

### 【实训目标】

1. 深入理解和熟练掌握统计调查知识和调查方法。
2. 加强理论和实践的联系，提高将统计方法应用于日常学习与生活的能力。

### 【具体要求】

利用业余时间组织一次在疫情期间出租车司机的收支情况调查活动，设计调查目的、调查对象和调查单位、调查项目、调查时间和工作期限、调查经费预算、调查组织实施计划等。

### 【实训记录】

### 【交流与评价】

| 项 目  | 成 绩               |                   |                   |     |
|------|-------------------|-------------------|-------------------|-----|
|      | 收集资料的真实性<br>(30%) | 收集资料的全面性<br>(30%) | 调查结果的合理性<br>(40%) | 总 分 |
| 自我评价 |                   |                   |                   |     |
| 同学评价 |                   |                   |                   |     |
| 教师评价 |                   |                   |                   |     |







## 项目三 熟悉统计整理

### 知识目标 →

- 了解统计整理的意义、要求和工作步骤，了解统计数据的审核、订正技术
- 理解统计分组的概念和作用，掌握统计分组方法，理解统计分组体系
- 理解频数分布的概念和分类，掌握变量数列的编制
- 理解统计表的概念、构成要素，了解统计表的分类和设计，了解统计图的绘制方法

### 能力目标 →

- 能够利用统计整理的方法进行统计
- 可以运用统计分组进行统计
- 掌握统计图的绘制方法

### 思维导图 →

#### 熟悉统计整理

##### 认识统计整理

统计整理的意义和步骤

统计数据的审核

##### 理解统计分组

统计分组的含义、作用和基本类型

统计分组的方法和体系

##### 掌握频数分布

频数分布的含义和种类

变量数列的编制和图形显示法

##### 了解统计汇总与显示

统计汇总和统计表

统计图



## 项目情境 →

新型冠状病毒肺炎疫情对国内多个行业造成了较大的冲击，其中包括房地产行业。据统计，2022年1~7月，全国房地产开发投资79 462亿元，同比下降6.4%；其中，住宅投资60 238亿元，同比下降5.8%。

2022年1~7月，房地产开发企业房屋施工面积859 194万平方米，同比下降3.7%。其中，住宅施工面积607 029万平方米，同比下降3.8%。房屋新开工面积76 067万平方米，同比下降36.1%。其中，住宅新开工面积55 919万平方米，同比下降36.8%。房屋竣工面积32 028万平方米，同比下降23.3%。其中，住宅竣工面积23 279万平方米，同比下降22.7%。

通过上述内容，对于房地产行业现状的描述属于统计整理吗？统计整理的方法有哪些？

---

---

---

---

---

---

---

---

---

---



## 任务一 认识统计整理

### 任务卡 1

某学校进行期中考试，面对上千名学生的成绩，如何公布才能使家长一目了然呢？假如你是该学校校长，你将如何统计和公布成绩？

### 知识卡 1 统计整理的意义和步骤

#### 一、统计整理的意义

统计整理是根据统计研究的目的和任务，对所收集到的多项原始资料进行科学的分类、汇总和加工，使之系统化、条理化的工作过程。统计整理既包括对统计调查所得到的原始资料进行整理，也包括对加工过的综合资料，即次级资料进行再整理。

统计整理的意义如下：

（1）通过对统计资料的加工整理，使之系统化、条理化，可以通过综合指标对总体做出概括性的说明，可以揭示出总体的内在特征。

（2）通过统计调查所收集到的资料，只有经过科学的审核、分类、汇总等整理工作，才能使统计在认识社会的过程中，实现由个别到全体、由特殊到一般、由现象到本质、由感性到理性的转化，才能从整体上反映出事物的数量特征。

（3）统计研究中经常要进行动态分析，这就需要有长期积累的历史资料，而根据积累资料的要求，要对已有的统计资料进行筛选，以及按历史的口径对现有的统计资料进行重新调整、分类和汇总等，都必须通过统计整理工作来完成。



## 二、统计整理的步骤

统计整理是一项十分复杂而细致的工作，要做好这项工作，必须有组织、有计划地采用科学的方法。统计整理通常按下列步骤进行：

（1）设计方案。设计方案即设计统计汇总方案，明确规定统计分组的方法和设置汇总的统计指标。

（2）资料审核。资料审核即主要审核原始资料的准确性、完整性、及时性和适用性。

（3）统计分组。统计分组是指采用科学的方法，对审核后的原始资料进行分组。

（4）统计汇总。统计汇总是指对统计分组后的资料进行汇总和计算，计算出各组指标和综合指标。

（5）编制统计表，绘制统计图。

上述五个步骤中，设计方案和资料审核是统计整理的前提，统计分组是统计整理的关键，统计汇总是统计整理的中心，统计表、统计图是统计整理的成果显示。根据统计研究的目的和任务以及统计对象的特点不同，统计分组的方法也不同。



想一想

### 任务卡 2

请根据自己的理解，勾选出统计数据审核的内容。

☐ 完整性审核

☐ 适用性审核

☐ 快速性审核

☐ 准确性审核

☐ 时效性审核

☐ 精准性审核

---

---



## 知识卡 2 统计数据的审核

### 1. 数据审核的内容

统计数据的审核是保证统计数据整体质量的重要手段，它包括汇总前的审核和汇总后的审核。数据审核的内容包括以下几点：

（1）完整性审核。完整性审核主要审核应调查的单位或个体是否有遗漏，所有的调查项目或指标是否填写齐全。

（2）准确性审核。准确性审核主要审核数据是否真实反映客观实际情况，内容是否符合实际，审核数据是否有错误，计算是否正确等。

（3）适用性审核。适用性审核主要审核数据的来源、数据的口径以及相关的背景材料，审核这些数据是否符合分析研究的需要。

（4）时效性审核。有些统计数据时效性很强，如果滞后，则会失去研究意义，因此，对于这类数据要进行时效性审核，以确保使用最新的统计数据。

此外，对汇总后的统计资料也要进行审核，以保证统计汇总资料的准确性。对汇总后的统计资料的审核主要从复计审查、对照审查和表实审查等几个方面进行。

### 2. 数据审核的方法

数据审核的方法包括逻辑检查和计算检查。

（1）逻辑检查是从定性角度，审核数据是否符合逻辑，内容是否合理，各项目或数字之间有无相互矛盾的现象。逻辑检查主要用于对定类数据和定序数据的审核。

（2）计算检查是检查调查表中的各项数据在计算结果和计算方法上有无错误。计算检查主要用于对定距数据和定比数据的审核。

### 3. 数据审核的订正

在数据审核过程中发现有差错的地方，应根据不同情况分别进行处理：发现有缺报、缺份或缺项等情况，应及时催报、补报；对于可以肯定的错误要及时更正并通知相关单位；对于不能确定但存有疑问的数据要进行复核，查出的错误应及时更正；对于错误现象较多的情况，应该重新填报。填报虚假数据，一经发现要严肃处理，触犯法律的要移交司法部门。





## 任务二 理解统计分组

### 任务卡 1

某学校在开学当天将重新调整学生的座位，现征求学生的意见，有的建议以成绩高低进行排位，有的建议以身高行排位，有的建议以性别分组来排位。假如你是该学校的学生，你想如何进行排位？请说明理由。

### 知识卡 1 统计分组的含义、作用和基本类型

#### 一、统计分组的含义

统计分组是指根据统计总体内在的特征和统计研究的任务需要，将统计总体按照选定的变异标志划分为若干组成部分的一种统计方法。统计资料分组的目的，就在于把统计总体中的具有不同性质的单位分开，把性质相同的单位合并在一起，保持各组内统计资料的一致性和各组之间统计资料的差异性，以便进一步研究调查对象的数量表现与数量关系，进而正确认识调查对象的本质及规律性。例如，在我国人口普查中，作为个体的每个人，在年龄、性别、民族、文化程度以及居住地等诸多调查标志上不完全相同。为反映我国人口总体内部的差异，就需要按照不同的标志对全国人口进行分组。例如，按性别可分为男、女两组，按年龄、民族可划分为若干组，这有助于对我国人口的性别、年龄、民族等各方面的结构及其比例关系的认识。

#### 二、统计分组的作用

统计分组的作用主要表现为以下几点。



拓展窗

### 1. 区分总体类型

现象的类型是多种多样的,不同类型的现象存在本质差别,通过统计资料的分组就可以把不同类型的现象区别开来。

### 2. 反映总体内部结构

通过分组,统计总体被划分为若干组成部分,计算各组成部分的总量在总体总量中所占的比重,即可反映总体结构特征与总体结构类型。

### 3. 分析总体在数量现象之间的依存关系

现象之间总是相互联系、相互依存、相互制约的,分组就是要在现象各种错综复杂的联系中,找出它们内在的联系和数量关系。具体做法:将一个可变标志(自变量)作为分组标志,观察另一个标志(因变量)相应的变动状况,如居民家庭收入与就业人数有密切的联系,通过分组就可以反映这两个标志之间相互联系的程度和方向。

## 三、统计分组的基本类型

根据统计研究的目的和任务以及统计对象的特点不同,统计分组的方法也不同。

### 1. 按照分组标志的性质不同分类

按照分组标志的性质不同,可将统计分组分为品质标志分组和数量标志分组。

(1) 按品质标志分组是指用品质标志将总体分为若干组来表示。按品质标志分组,标志一经确定,组数和组限都非常明显。这些组在组限上、性质上都很明显,不存在组与组之间界限难以划分的问题。在统计工作中,对于比较复杂的分组则由上级统计机关或各业务主管部门统一编制标准的分类目录供人们使用。

(2) 按数量标志分组就是选择反映社会经济现象数量差异的数量标志作为分组标志。例如,职工按工资收入多少分组,企业按产值计划完成程度分组等。

### 2. 按照分组标志的多少分类

按照分组标志的多少不同,可将统计分组分为简单分组和复合分组。

(1) 简单分组是指按一个标志进行分组,如表 3-1 所示(试统计全班学生统计课程考试成绩并分组,填入表中)。



表 3-1 ×× 班学生统计课程考试成绩分组

| 成 绩     | 人数 / 人 |
|---------|--------|
| 60 分以下  |        |
| 60~70 分 |        |
| 70~80 分 |        |
| 80~90 分 |        |
| 90 分以上  |        |
| 合计      |        |

(2) 复合分组是指对同一个总体在同一张分组表上, 把两个或两个以上标志层叠起来进行分组, 如表 3-2 所示。

表 3-2 ×× 市食品企业按所有制和固定资产原值的复合分组

| 按所有制<br>分组 | 按固定资产原值分组 /<br>万元 (不含上限) | 企业数<br>/ 个 | 职工人数<br>/ 人 | 总产值<br>/ 万元 | 人均产值 /<br>(万元·人 <sup>-1</sup> ) |
|------------|--------------------------|------------|-------------|-------------|---------------------------------|
| (甲)        | (乙)                      | (1)        | (2)         | (3)         | (4)=(3)<br>/(2)                 |
| 国有企业       | 200 以下                   |            |             |             |                                 |
|            | 200~350                  |            |             |             |                                 |
|            | 350~500                  |            |             |             |                                 |
|            | 小计                       |            |             |             |                                 |
| 私营企业       | 200 以下                   |            |             |             |                                 |
|            | 200~350                  |            |             |             |                                 |
|            | 350~500                  |            |             |             |                                 |
|            | 小计                       |            |             |             |                                 |
|            |                          |            |             |             |                                 |



### 任务卡 2

QQ 和微信是我们现在常用的通信工具，你是按什么对好友进行分组的？

## 知识卡 2 统计分组的方法和体系

### 一、统计分组的方法

统计分组的关键在于正确选择分组标志和划分各组界限。

#### 1. 正确选择分组标志

进行统计分组，需要有分组的依据。确定分组的依据就是选择分组标志，它是分组时划分组别的标准。划分各组界限，就是要在分组标志的变异范围内，划定各相邻组间的性质界限和数量界限。

任何事物都有许多标志，选择不同的分组标志会得出不同的结果。标志选择不当，分组结果就不能正确反映总体的性质特征。因此，正确地选择分组标志是科学分组的前提，也是统计资料整理的关键。正确选择分组标志，需要注意以下几点：

(1) 根据统计研究的目的选择分组标志。统计总体的各个单位有许多标志，选择什么标志作为分组标志，要根据统计研究的目的而定。例如，在某高等院校在校学生这一总体中，每一个在校学生是总体单位。学生有年龄、身高、性别、体重、民族、政治面貌和学习成绩等许多标志。如果要分析该校学生的年龄结构，就要选择年龄作为分组标志；如果要反映学生的学习成绩构成，就要选择每门课程的平均成绩作为分组标志；如果要研究学生的性别结构，就要选择性别作为分组标志。可见，对于不同的研究目的，需要选择不同的分组标志。

(2) 选择最能反映事物本质特征的标志进行分组。事物的标志多种多样，有些标志是带有根本性的、主要的标志，能够反映事物的本质，而有些则是非



本质的次要标志。例如，要研究我国经济结构的特点，经济类型、产业结构等就是最基本的标志。再如，要研究企业的经济效益好坏，可供选择的标志也很多，如总产值、净产值、增加值、销售收入、利税额、劳动生产率、单位产品成本和资金占用额等，但更能综合反映企业经济效益的是利税额。

因此，在进行分组时，要从统计研究的目的出发，从若干标志中选择最能反映事物本质特征的主要标志。

(3) 考虑现象发展的历史条件和经济条件。社会经济现象在不断发展变化，历史条件变了，事物的特征也会发生变化，也就是说，最能反映现象本质特征的标志将随之改变。例如，在研究中国的人口状况时，阶级是反映人口本质特征的标志，人口按阶级分组就是最本质、最重要的分组。在不同的经济条件下，分组标志的选择也是不同的。例如，对于劳动密集型产业，应选择职工人数作为分组标志来反映各企业生产规模的大小；对于技术密集型产业，就要选择固定资产价值或产值作为分组标志来反映各企业生产规模的大小。

### 2. 按品质标志或按数量标志分组

按分组标志的特征不同，统计总体可以按品质标志分组，也可以按数量标志分组。

(1) 按品质标志分组。按品质标志分组是指选择反映事物属性差异的标志作为分组标志，并在品质标志的变异范围内划定各组界限，将总体划分为若干性质不同的组成部分。例如，在研究国民经济总体时，可以通过按经济类型、隶属关系、地区、国民经济部门等品质标志分组。按品质标志分组有时界限明确，比较容易，有时却比较复杂。

在实际工作中，常常需要对所研究的现象进行复杂的品质分组。这种复杂的品质分组也称分类，如国民经济行业分类、产品分类和人口职业分类等。它们不仅涉及复杂的分组技术，而且涉及国家的政策和有关科学理论，因而在分组时要十分慎重。为了保证各种分类的统一性和完整性，国家制定了统一的分类目录和标准。

(2) 按数量标志分组。按数量标志分组是指选择反映事物数量差异的数量标志作为分组标志，并在数量标志的变异范围内划定各组界限，将统计总体划分为性质不同的若干组成部分，如企业按固定资产价值分组、人口按年龄分组等。



与品质标志不同,数量标志具体表现为许多不等的变量值,这些变量值能准确反映现象数量上的差异,却不能明确反映现象性质上的区别。因此,在按数量标志进行统计分组时,应根据研究的目的,首先确定总体在已选定的数量标志的特征下有多少种性质不同的组成部分,然后研究确定各组成部分的数量界限,使分组的数量界限能够区分现象性质上的差别。

### 3. 简单分组和复合分组

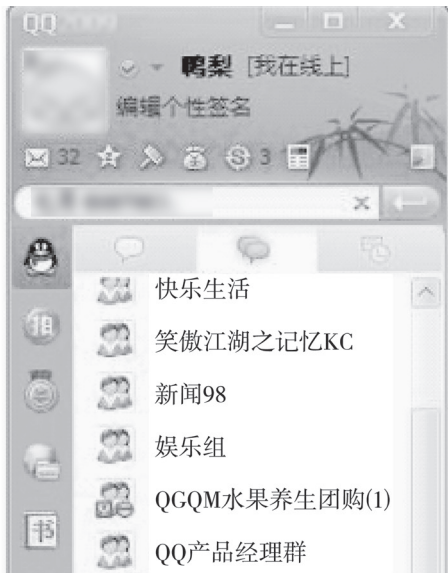
按照选择分组标志多少的不同,可以将统计分组分为简单分组和复合分组。

(1) 简单分组。简单分组是指对所研究的总体按一个标志进行分组。例如,工业企业按生产规模分为大型、中型和小型三个组,人口按性别分为男性、女性两个组,投资按资金来源分为国家预算内资金、国内贷款、利用外资、自筹投资和其他投资五个组。这种分组比较简单,它只能说明现象在某一方面的分配状况或联系。

简单分组其实经常出现在我们身边,图 3-1 为 QQ 软件的好友分组和群分组。



(a) 好友分组



(b) 群分组

图 3-1 QQ 软件的好友分组和群分组

(2) 复合分组。复合分组是指对所研究的总体按两个或两个以上的标志进行的多层次分组。例如,工业企业按经济类型分组后,每一组中再按规模进行分组;人口按性别分组后再按年龄分组等。采用复合分组能更深入地反映总体





的内部结构，更细致地分析问题。但是，随着分组标志的增加，组数将成倍增加，因此也不宜采用过多的标志进行复合分组。究竟采用几个标志进行复合分组，要根据统计研究的目的和任务来决定。

## 二、统计分组体系

统计分组体系分为平行分组体系和复合分组体系。

### 1. 平行分组体系

平行分组体系又称并列分组体系。将社会经济现象总体只选择一个标志分组称为简单分组。对同一总体选择两个或两个以上的标志分别进行简单分组，排列起来，即成为平行分组体系。例如，我国企业分别按经济类型、隶属关系分组，形成两个不同角度的观察结果，如图 3-2 和图 3-3 所示。

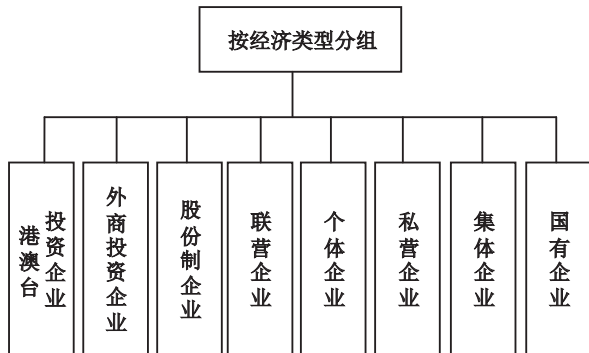


图 3-2 按经济类型划分我国企业

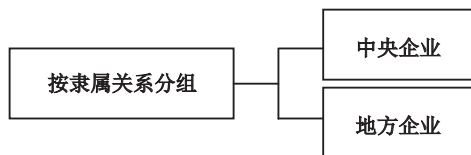


图 3-3 按隶属关系划分我国企业

### 2. 复合分组体系

复合分组体系又称交叉分组体系，它是用两个或两个以上的分组标志重叠起来对总体进行的分组。多个复合分组组成的体系就形成了复合分组体系，如按轻重工业和企业规模对我国企业进行分组（见图 3-4）。

复合分组体系的特点是两种或多种分组相互独立而不重叠，既可从不同的方面反映事物的多种结构，又不致使分组过于烦琐，故被广泛采用。



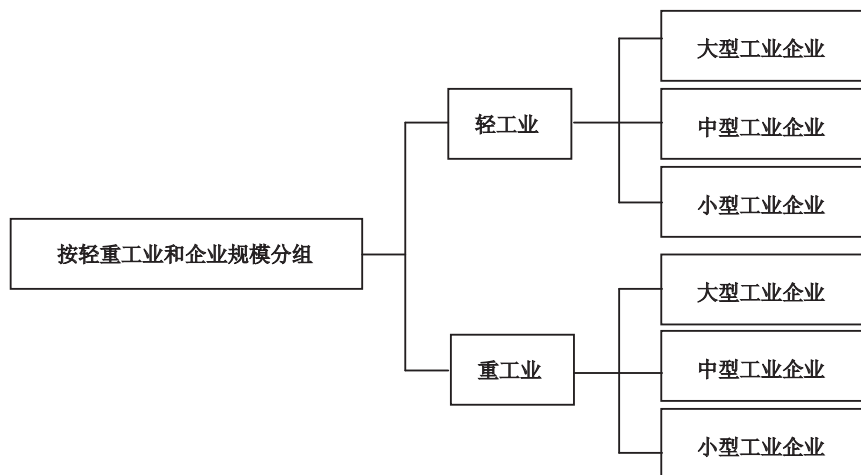


图 3-4 按轻重工业和企业规模划分我国企业

### 任务三 掌握频数分布

#### 任务卡 1

大家都听说过古代流传下来的“三六九等”的说法，这句话是什么意思？它和统计分组有联系吗？有怎样的联系？

#### 知识卡 1 频数分布的含义和种类

##### 一、频数分布的含义

频数分布是指在统计分组的基础上，把总体全部单位按组归类整理，汇总出各组的总体单位数，并将其按分组顺序加以排列而形成的统计数列。分布在各组中的总体单位数，称为频数，亦称次数；频数与总频数的比值，称为比率，



亦称频率。频数分布是对统计分组结果的表达,是进一步统计分析的重要基础。

## 二、频数分布的种类

性质不同的现象呈现不同的频数分布,频数分布大致可以分为以下三种类型。

### 1. 钟形分布

钟形分布的特征是靠近中间的变量值分布的次数多,靠近两端的变量值分布的次数少,即“两头小,中间大”。如果将变量值与其对应的频数在直角坐标系中对应的点连接起来绘制成曲线图,宛如一口钟,所以称为钟形分布。

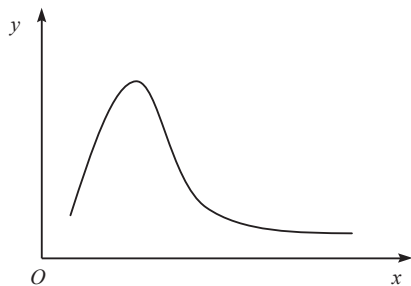
在自然或社会经济现象中,有许多频数分布是钟形分布。例如,农作物的单位面积产量、零件的公差、纤维强度等都属于钟形分布。

根据两侧的频数是否均衡,钟形分布可分为对称分布与非对称分布。对称分布又称正态分布,它以变量值的中点为对称轴,两侧变量值分布的次数随着离变量值中点距离的增大而逐渐减小,减小的频数基本对等,如图 3-5 所示。

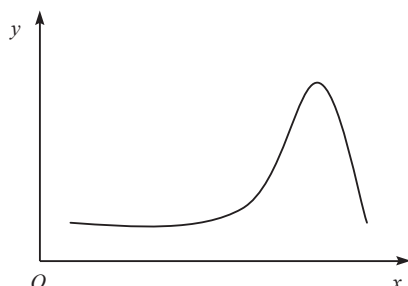


图 3-5 对称分布(正态分布)

非对称分布又称偏态分布。根据不同方向和程度的偏斜,偏态分布可分为图形向右偏斜的右偏分布和图形向左偏斜的左偏分布,如图 3-6 所示。



(a) 右偏分布



(b) 左偏分布

图 3-6 非对称分布(偏态分布)



## 2. U 形分布

U 形分布的特征是两端的频数分布多, 中间的频数分布少, 形成“两头大, 中间小”的频数分布特征, 其频数分布曲线类似字母 U, 如图 3-7 所示。例如, 人和动物的死亡率分布呈 U 形分布。

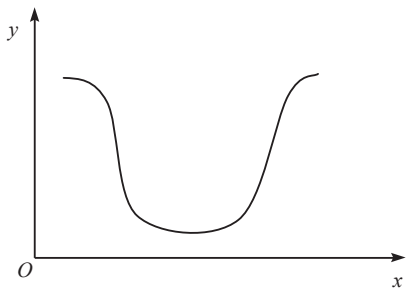


图 3-7 U 形分布

## 3. J 形分布

J 形分布的特征是“一边小, 一边大”, 即大部分变量值集中在某一端分布, 有正 J 形分布和反 J 形分布两种, 如图 3-8 所示。正 J 形分布是频数随着变量值的增大而增多。例如, 经济学中的供给曲线, 随着价格的提高供给量以更快的速度增加, 呈正 J 形分布。反 J 形分布是频数随着变量值的增大而减小。例如, 经济学中的需求曲线, 随着价格的提高需求量以较快的速度减少, 呈反 J 形分布。

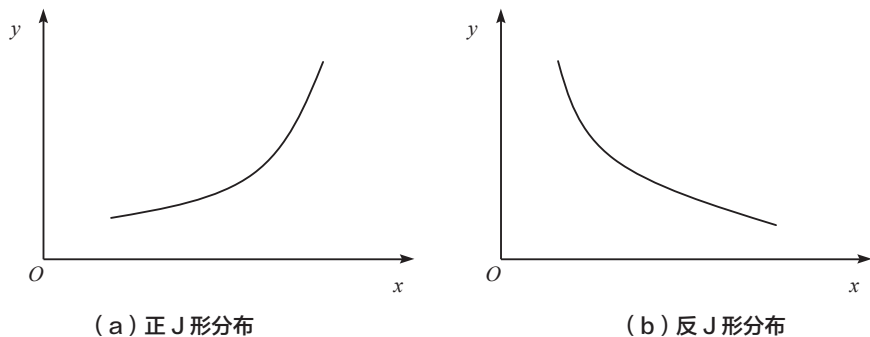


图 3-8 J 形分布

### 任务卡 2

请你列举出你所知道的常用的变量分布图形。

☐ 直方图

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_



## 知识卡 2 变量数列的编制和图形显示法

### 一、变量数列的编制

变量数列的编制应按以下步骤进行：

(1) 将原始资料按数值大小依次排列。

(2) 确定组数和组距。组数的多少和组距的大小是相互制约的。组数越多，组距就越小；反之，组数越少，组距就越大。等距数列的组距等于全距除以组数。对于组数和组距，先确定哪一个，不能机械地规定，而要根据实际情况来定。确定组距和组数应考虑下列原则：

①尽量能反映出总体单位的分布情况及总体单位的集中趋势。

②尽可能区分出组与组性质上的差异。

(3) 确定组限。组限的确定主要考虑下列几点：

①最小组的下限要略低于最小变量值，最大组的上限要略高于最大变量值。

②组限的确定应当有利于表现总体单位分布的规律性。

③对于等距数列，如果组距是 5, 10, 15, …, 100, …则每组的下限最好是它们的倍数。

由于变量有连续型和离散型之分，其组限的划分要求也不同。对于连续型变量（如年龄、身高等），划分组限时，相邻组的组限必须重合。对越高越好的现象，习惯上规定，各组不包括其上限变量值的单位，即所谓“上组限不在内”原则。对越低越好的现象，习惯上规定，各组包括其上限变量值的单位，即所谓“上组限在内”原则。对于离散型变量（如企业数、零件数等），划分组限时，相邻组的组限必须间断。但是，应当指出，在实际工作中，为了保证不重复、不遗漏总体单位，对于离散型变量也常常采用连续型变量的组限表示方法。

### 二、变量数列图形显示法

常用的变量分布图形有直方图、折线图和圆滑曲线图。

#### 1. 直方图

直方图是指在横纵轴之间以直方条形来显示频数分布的图形。对于等距数列，左纵轴表示频数，右纵轴表示频率，横轴表示变量值。频数和频率的显示



应与实际资料的计算相符。例如，总体单位总数（总频数）为 40，左纵轴的 40 应与右纵轴的 100% 对应，这时 20 应与 50% 对应。不同的频数要与相应的频率一一对应。

表 3-3 为某班 40 名学生的学习成绩分组。根据表 3-4 的等距数列可画出直方图，如图 3-9 所示。

表 3-3 某班 40 名学生的学习成绩分组

| 成 绩      | 人数 / 人 |
|----------|--------|
| 50~60 分  | 2      |
| 60~70 分  | 4      |
| 70~80 分  | 11     |
| 80~90 分  | 18     |
| 90~100 分 | 5      |
| 合计       | 40     |

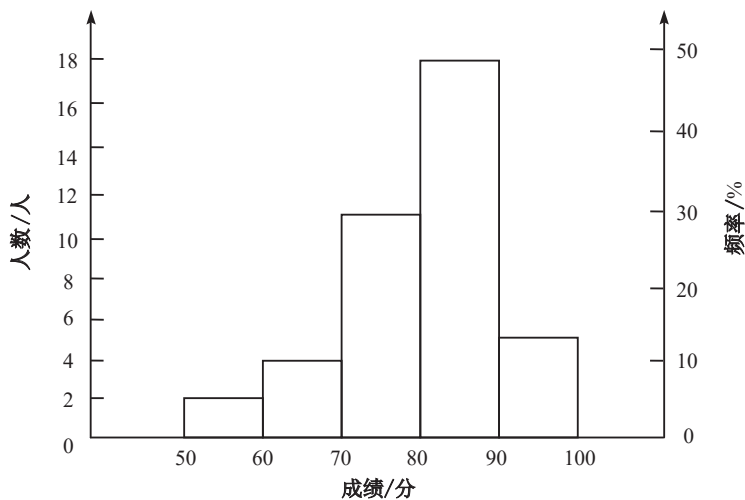


图 3-9 直 方 图

## 2. 折线图

将直方图各条形顶端中点两两连接起来，所形成的图形称为折线图。它是在直方图的基础上形成的，如图 3-10 所示。

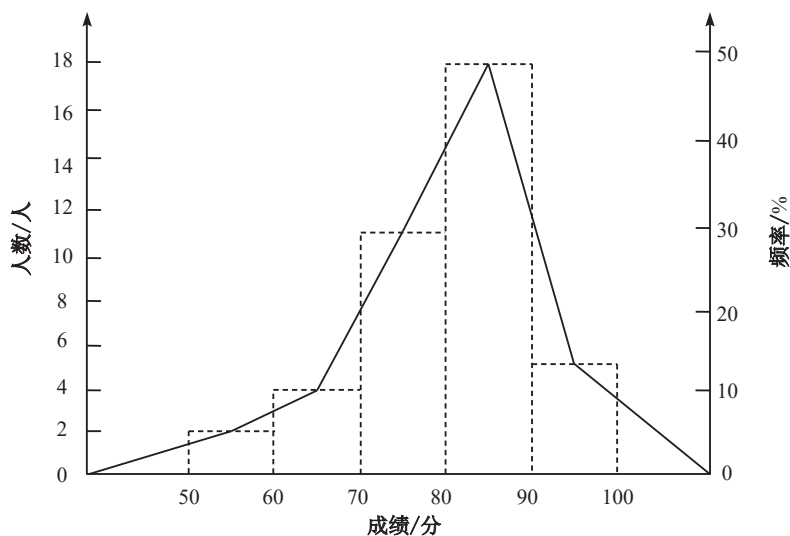


图 3-10 折线图

### 3. 圆滑曲线图

当各组组距无限缩小时，折线图相邻条形的中点将无限接近，折线图的折线变圆滑，进而成为曲线。

圆滑曲线图是用曲线显示频数分布情况的图形。一般用等距分组资料绘制条形图，再按绘制折线图的方法将连线用平滑曲线表示出来，就形成了频数分布曲线图，如图 3-11 所示。

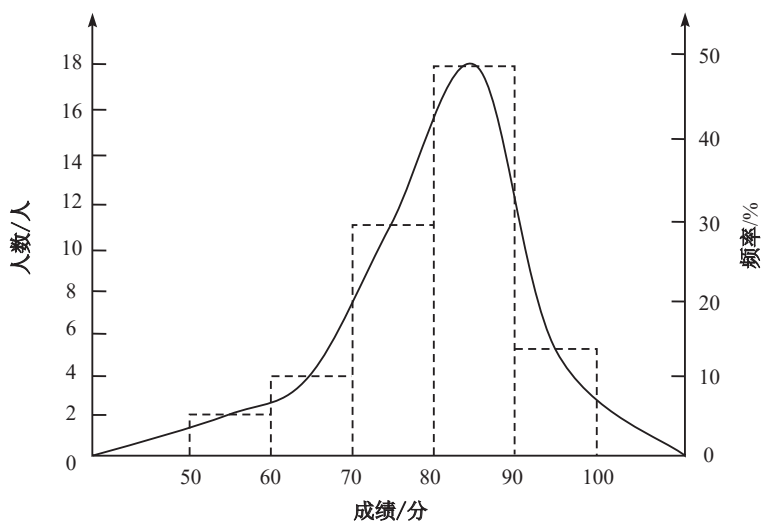


图 3-11 频数分布曲线图



## 任务四 了解统计汇总与显示

### 任务卡 1

你对统计表了解吗？请查找资料并列举出在生活中运用到统计表的例子。

### 知识卡 1 统计汇总和统计表

#### 一、统计汇总

在统计分组的基础上，将总体各单位和部分各单位标志值分别归纳到各组，然后计算出各组总的合计数，通过汇总便可得到反映总体特征的各种统计指标的过程，称为统计汇总。统计汇总也是统计整理的重要内容。

##### （一）统计汇总的内容

按照标志的不同特点，统计汇总的内容可以分为以下三点。

##### 1. 总体单位总量方面的汇总

总体单位总量方面的汇总，或者称次数、频数的汇总，即汇总各组 and 总体的单位总个数。这一内容的汇总结果，就是总体单位数量。它是研究总体在分组标志上的一般分布状况的直接依据和基础，也是进一步深入分析的重要依据。

##### 2. 绝对数标志值的汇总

绝对数标志值的汇总即绝对数或总量形式的标志值在各组的加总，最终合计为指标即总体标志总量。

##### 3. 平均数和相对数标志值的汇总

按照标志的严格定义，总体单位的数量标志中包括平均数和相对数。总体





各单位的平均数和相对数标志值不能直接相加，因此，在汇总这两类标志时必须先把它们的绝对数分子、分母分解出来，然后把绝对数分子、分母进行加总，通过对加总结果的对比计算，汇总总体各单位在各组的平均数和相对数标志值。

### （二）统计汇总的组织形式

统计汇总是一项十分复杂的工作，需要有一整套科学的组织形式，以保证统计汇总的顺利进行。统计汇总一般有逐级汇总、集中汇总和综合汇总三种组织形式。

#### 1. 逐级汇总

逐级汇总是指自下而上一级一级地汇总本地区、本系统或本单位的调查资料。这种组织形式能够满足各级单位对资料的需要，也便于就地审核和订正原始资料，但汇总层次多，比较费工费时，出现差错的概率也较大。

#### 2. 集中汇总

集中汇总是把全部调查资料集中在一个机关或者最高统计机关进行一次汇总。这种组织形式可以缩短汇总时间，减少汇总差错，但原始资料如有差错不能就地更正，整理结果有时不能及时满足各地区、各部门的需要。

#### 3. 综合汇总

综合汇总是将逐级汇总和集中汇总结合起来的一种组织形式，即对于各地区和各级都需要的基本资料实行逐级汇总，对需要在全国或本系统范围内进行加工的资料则实行集中汇总。

近年来，计算机技术飞速发展，人们借助于现代计算机技术和网络技术可以实现各级统计汇总的同步进行，逐级汇总、集中汇总、综合汇总的含义也发生了变化。例如，采用逐级汇总这种组织形式进行资料汇总时，只要各基层单位将原始数据输入数据库，各级汇总工作即可同时展开，上级部门的汇总工作不再需要等下一级的汇总工作结束之后才能进行，统计汇总的效率和质量将会大大提高。

### （三）统计汇总的审核

统计汇总的审核包括汇总前的审核和汇总后的审核两个环节。汇总前的审核是把握统计汇总质量的关键，审核的主要内容是资料的准确性、及时性和完整性。汇总后的审核是检查汇总工作的质量，审核的主要内容是汇总结果的真



实性和准确性。

#### （四）统计汇总的方法

统计汇总主要有手工汇总法和机械汇总法两种。

##### 1. 手工汇总法

手工汇总法也称手工整理法，它是用手工方法利用一些简单工具对统计资料进行整理，其整理速度慢、时效性差，也比较容易出现差错，仅在资料较少、人力又许可的条件下采用。常用的手工汇总法有划记法、过录法、折叠法和卡片法四种。

（1）划记法。划记法是按照事先分好的组用点线符号（如“正”字）计算各组的单位数和合计总数。这种方法简便易行，但容易出现错漏，也不能汇总各组 and 总体单位的标志值，一般在总体单位资料不多的情况下采用。

（2）过录法。过录法是将调查资料先过录到事先设计好的整理表上，并计算出各种合计数，然后将结果填入正式的统计汇总表中。这种方法汇总的内容比较全面，也便于校对检查，但工作量大，费时费力。

（3）折叠法。折叠法是将所有调查表中需要汇总的项目和数值折在边上，一张接一张地叠在一起直接汇总同一纵栏或同一横行中的数字。这种方法简便易行，但汇总时必须细致，应随时进行检查。

（4）卡片法。卡片法是将每个总体单位需要汇总的项目和数值摘录到事先准备好的卡片上，然后根据卡片进行分组和汇总计算。在调查资料多、统计分组细的情况下，宜采用卡片法进行汇总，它比划记法、过录法和折叠法的汇总质量要高。

##### 2. 机械汇总法

机械汇总法也称机械整理法，它是利用机器设备对统计资料进行整理。对于大量的统计资料，靠手工的办法难以迅速、准确、及时地加以处理。手工汇总法在实际应用中受到极大的限制，机械汇总法的应用越来越广泛。随着科学技术的不断发展，统计资料整理经历了算盘、计算尺、手摇计算机、电动计算机到电子计算机汇总的历程，其汇总速度越来越快，汇总质量越来越高。电子计算机具有强大的数据处理功能，而且具有运算速度快、精度高，数据存储量大，修改、存储、取用方便，程序自动执行等特点，这些都是手工汇总法无法



比拟的。

近年来,微型电子计算机得到了广泛的应用,它把人们从大量繁杂的手工数据处理工作中解脱出来,而且完成了许多用手工难以处理的工作,极大地提高了工作效率和工作质量。一般来说,电子计算机对数据进行处理,包括对数据的收集、记载、修改、分类、排序、检索、存储、计算、传输、制表等工作。应用电子计算机进行统计资料的数据处理(或汇总)通常采用如下步骤:

(1)明确所要编制程序的目的。明确所要编制程序的目的即确定程序要完成什么功能,它完全取决于统计汇总的目的和要求。

(2)进行技术准备工作。进行技术准备工作是指要掌握准备采用的电子计算机语言。

(3)进行可行性分析。可行性分析的目的是取得一个技术上可行、经济上合理、实施起来有效的数据处理方案。这就需要根据现有的基础、技术、人力等环境条件,结合费用和效果等具体分析所使用电子计算机本身的软件和硬件功能是否能满足所要进行数据处理(或汇总)的目的和要求。

(4)根据汇总方案画出程序框图。这是编写程序的重要步骤,框图应力求明了。

(5)进行代码设计。进行代码设计是指对事物进行编码,如把汉字信息数字化。它的作用是使数据标准化、系统化,从而实现对数据资源的有效处理。代码设计是一项复杂的工作,其质量如何不仅影响数据输入的速度和质量,还影响数据处理的最终结果。因此,在设计代码时要十分谨慎,它往往需要花费很长的时间(如全国的商品编码一般要用十几年的时间才能完成),编码体系一旦确定便不宜再修改。

(6)编写程序。设计程序时应尽量使程序结构模块化。这种程序便于阅读,结构清楚,条理性强,容易发现并修改错误。

(7)程序调试及运行。程序调试包括纠正程序中的各种错误,对程序的总体布局进行优化、改进等。最后,由电子计算机执行程序进行统计制表,并通过输出设备将结果打印出来。

必须指出,现代计算机技术发展迅猛,计算机应用技术也是日新月异,这就为利用计算机进行资料整理提供了极大的方便。人们甚至可以不经上述七个步骤,只需安装好某些专用软件就可以直接在计算机上进行各种汇总工作和



统计计算与分析工作。

统计资料整理要实现现代化,就要运用先进的统计科学和现代计算机技术。利用现代计算机技术、数据传输技术和网络技术,建立以计算机网络为主要特征的计算中心和信息交流中心,是统计资料整理工作的发展方向。计算机网络是指将地理位置不同,并具有独立功能的多个计算机系统通过通信线路和通信设备连接起来,且以功能完善的网络软件实现资源共享的系统。这种系统可以使网络内的所有用户都能享用网上各计算机系统的全部或部分资源。它能够克服地理条件的限制,把分散在不同地区、不同行业 and 不同部门的计算机联成网络整体,实现数据和信息资源的共享,使手工方式逐级汇总时失掉的大量数据和信息资源得到开发和利用,即使在原始数据不变的情况下也能使可供信息量成倍增长。同时,这种现代化汇总系统可以避免系统中的重复劳动和投资,也便于均衡系统负载,提高系统的处理能力和可靠性,大大加快资料整理的速度和数据传输速度,并提高汇总数据的准确性。

## 二、统计表

### (一) 统计表的结构

通过统计汇总得出表明社会经济现象总体单位数和一系列标志总量的资料,把这些资料按一定顺序在表格上表现出来,这种表格称为统计表。

这里,我们是把统计表当作整理过程的最后一个环节来看的。汇总的结果体现在表上,意味着整理过程的终止。但是,统计表应从广义方面来看,任何用以反映统计资料的表格都是统计表。

数字是统计的语言。统计研究社会经济现象的数量关系,主要通过数字资料来表现。统计表和统计图都是系统地表述数字资料的基本形式。统计表能够系统地组织和合理安排大量数字资料,便于对照比较,使得统计资料的表现显得紧凑、有力、突出,因而在描述统计资料中得到广泛的运用。

从外表形式看,统计表是由纵横线交叉的一种表格所组成的,在表格上填写着反映社会经济现象的数字资料。因此,统计表由标题、横行和纵栏、数字资料等部分构成。标题分为三种:总标题是表的名称,放在表的顶端中央;横行标题或称横标目,写在表的左方,纵栏标题或称纵标目,写在表的上方,分别说明横行或纵栏所填列数字资料的内容。



统计表的内容包括主词和宾词两个部分。主词就是统计表所要说明的总体、总体的各个组或各个单位的名称。宾词是用来说明主词的各种指标的。通常情况下主词列在表的左方，即列于横行；宾词列在表的上方，即列于纵栏。但是，当这样排列会使统计表的表式过分狭长或过分宽短时，也可以将主词与宾词合并排列或变换位置排列。

下面举一个一般统计表的例子，表的组成部分在表旁加以说明，如图 3-12 所示。

| 表 X-X 某年我国固定资产投资分地区情况 |          |       |          |
|-----------------------|----------|-------|----------|
| 地 区                   | 完成投资额/亿元 | 增长百分数 | 占全部投资的比重 |
| 东部地区                  | 12 188   | 17.5% | 62.7%    |
| 中部地区                  | 4 121    | 20.7% | 21.2%    |
| 西部地区                  | 2 387    | 21.2% | 12.3%    |

图 3-12 一般统计表的组成部分示意图

## （二）统计表的种类

### 1. 按主词的结构分类

统计表的种类可根据主词的结构来决定，按主词是否分组和分组的程度，统计表可分为简单表、简单分组表和复合表。

（1）简单表。简单表是主词未经任何分组的统计表。例如，主词由总体单位清单组成的一览表，主词由地区、国家、城市等目录组成的区域表，以及主词由时间顺序组成的编年表，等等。表 3-4 是简单表的一个例子。

（2）简单分组表。简单分组表是指表的主词按某一标志进行简单分组的统计表。利用简单分组表来解释现象不同类型的不同特征，研究总体的内部构成，分析现象之间的依存关系。表 3-5 是简单分组表的一个例子。



表 3-4 某地区 12 个工业企业劳动生产率和固定资产利用效益

| 企业  | 经济<br>类型 | 职工人数<br>/ 人 | 固定资<br>产原值<br>/ 万元 | 产值<br>/ 万元 | 人均固定资产<br>/ 百元 | 每百元固定<br>资产产值<br>/ 百元 | 人均产值<br>/ 百元 |
|-----|----------|-------------|--------------------|------------|----------------|-----------------------|--------------|
| (甲) | (乙)      | (1)         | (2)                | (3)        | (4)=(2)÷(1)    | (5)=(3)÷(2)           | (6)=(3)÷(1)  |
| 1   | 国有       | 540         | 459                | 963.9      | 85             | 210                   | 178.5        |
| 2   | 国有       | 500         | 360                | 864.0      | 72             | 240                   | 172.8        |
| 3   | 国有       | 480         | 384                | 844.8      | 80             | 220                   | 176.0        |
| 4   | 集体       | 420         | 336                | 621.6      | 80             | 185                   | 148.0        |
| 5   | 其他       | 400         | 288                | 518.4      | 72             | 180                   | 129.6        |
| 6   | 集体       | 360         | 270                | 445.5      | 75             | 165                   | 123.8        |
| 7   | 其他       | 360         | 198                | 277.2      | 55             | 140                   | 77.0         |
| 8   | 国有       | 350         | 238                | 368.9      | 68             | 155                   | 105.4        |
| 9   | 集体       | 340         | 221                | 309.4      | 65             | 140                   | 91.0         |
| 10  | 集体       | 250         | 160                | 192.0      | 64             | 120                   | 76.8         |
| 11  | 其他       | 240         | 144                | 165.6      | 60             | 115                   | 69.0         |
| 12  | 其他       | 200         | 116                | 110.2      | 58             | 95                    | 55.1         |
| 合计  |          | 4 440       | 3 174              | 5 681.5    | 71.5           | 179                   | 128.0        |

表 3-5 某地区工业企业按固定资产原值分组的劳动生产率和固定资产利用效益

| 按固定资产原值<br>分组 | 企业数<br>/ 个 | 职工人数   |        | 人均总产值<br>/ 百元 | 每百元固定资<br>产产值 / 元 |
|---------------|------------|--------|--------|---------------|-------------------|
|               |            | 人数 / 人 | 百分比    |               |                   |
| 200 万元以下      | 4          | 1 050  | 23.7%  | 70.9          | 121               |
| 200~350 万元    | 5          | 1 870  | 42.1%  | 121.0         | 167               |
| 350~500 万元    | 3          | 1 520  | 34.2%  | 175.8         | 222               |
| 合计            | 12         | 4 440  | 100.0% | 128.0         | 170               |

(3) 复合表。复合表是主词按两个或两个以上标志进行复合分组的统计表。在一定分析任务要求下, 复合表可以把更多的标志结合起来, 更深入地分析社会经济现象的特征和规律性。表 3-6 是复合表的一个例子。





表 3-6 某地区工业企业按经济类型和固定资产原值分组的劳动生产率和固定资产利用效益

| 按经济类型和固定资产原值<br>分组 |            | 企业数 / 个 | 人均固定资<br>产 / 百元 | 每百元固定<br>资产产值 / 元 | 人均产值<br>/ 百元 |
|--------------------|------------|---------|-----------------|-------------------|--------------|
| 国有                 | 200~350 万元 | 1       | 68.0            | 155.0             | 105.4        |
|                    | 350~500 万元 | 3       | 79.1            | 222.0             | 175.8        |
|                    | 小计         | 4       | 77.1            | 211.0             | 162.7        |
| 集体                 | 200 万元以下   | 1       | 64.0            | 120.0             | 76.8         |
|                    | 200~350 万元 | 3       | 73.8            | 166.4             | 122.9        |
|                    | 小计         | 4       | 72.0            | 158.9             | 114.5        |
| 其他                 | 200 万元以下   | 3       | 57.3            | 120.7             | 69.1         |
|                    | 200~350 万元 | 1       | 72.0            | 180.0             | 129.6        |
|                    | 小计         | 4       | 62.2            | 143.6             | 89.3         |
| 合计                 |            | 12      | 71.5            | 179.0             | 128.0        |

## 2. 按用途分类

统计表按用途可分为调查表、整理表和分析表。

(1) 调查表。调查表是指由统计调查的组织实施者根据统计调查的需要制定的用以对统计调查对象进行登记, 搜集相关原始数据和资料, 要求调查对象按照统一规定填报的表格。

(2) 整理表。整理表是指在统计整理中用于表现整理过程和结果的表格, 表中的数字是经过汇总的总量指标, 包括总体单位数和一系列标志总量。整理表也称为汇总表、综合表。

(3) 分析表。分析表是指在统计分析中用于整理所得的统计资料进行统计定量分析的表格。这类表格往往与整理表结合在一起, 成为整理表的延续。表中的数字既会有总量指标, 又会有在总量指标的基础上计算的多种相对指标和平均指标。

### (三) 宾词指标的分组配置

为使统计表的内容简明扼要, 应注重统计表宾词的指标配置。在宾词指标不要求进一步分组的情况下, 宾词配置就是指标体系的顺次列举, 根据指标说明问题的主次先后排列, 保持指标之间的逻辑关系。





当宾词指标需要分组时,宾词配置可有平行配置和层叠配置两种。平行配置就是宾词栏中各分组标志彼此分开,各标志的分组指标做平行排列;层叠配置则是将各分组指标层叠在一起,使各标志的分组指标有较大的增多。这样,在平行配置的情况下,宾词指标占有的栏数等于各标志的分组项数之和;而在层叠配置的情况下,宾词指标占有的栏数,要等于各标志的分组项数乘积。下面以某企业职工性别和工龄为例,列出宾词指标不同的分组配置表。平行配置如表 3-7 所示,层叠配置如表 3-8 所示。

表 3-7 某企业职工平行配置表

| 分 组 | 企业数<br>/ 个 | 职工人数<br>/ 人 | 性 别 |     | 工 龄   |        |        |
|-----|------------|-------------|-----|-----|-------|--------|--------|
|     |            |             | 男   | 女   | 5 年以下 | 5~10 年 | 10 年以上 |
| (甲) | (1)        | (2)         | (3) | (4) | (5)   | (6)    | (7)    |

表 3-8 某企业职工层叠配置表

| 分 组 | 企 业<br>数<br>/ 个 | 职工人数 / 人 |     |     | 工 龄   |     |     |        |     |      |        |      |      |
|-----|-----------------|----------|-----|-----|-------|-----|-----|--------|-----|------|--------|------|------|
|     |                 | 男        | 女   | 计   | 5 年以下 |     |     | 5~10 年 |     |      | 10 年以上 |      |      |
|     |                 |          |     |     | 男     | 女   | 计   | 男      | 女   | 计    | 男      | 女    | 计    |
| (甲) | (1)             | (2)      | (3) | (4) | (5)   | (6) | (7) | (8)    | (9) | (10) | (11)   | (12) | (13) |

从表 3-7 和表 3-8 可以看出,平行配置的宾词占有栏数为  $2+3=5$  栏,层叠配置的宾词占有栏数为  $2 \times 3=6$  栏。如果在宾词指标中再按年龄标志设置 5 个年龄组,这两种配置的差别就更大了。在平行配置的情况下,除原有栏数外,只要补充 5 栏,变为  $2+3+5=10$  栏。而在层叠配置的情况下,则要扩充到  $2 \times 3 \times 5=30$  栏。在这里还没有把合计栏计算在内。因此,对宾词指标的层叠配置要慎重考虑应用,它虽然能够详细说明研究对象的特征,但所用指标过多,会影响到统计表表现的明确性。

#### (四) 统计表的编制规则

统计表的编制,无论主词的内容或宾词指标的配置都要目的明确,内容鲜明,使读者能从表中看出研究现象的具体内容和情况。因此,在制表时,首先要强调目的要求,做到简明、紧凑、重点突出,避免过分烦琐。一个“包罗万象”的统计表,往往会使问题的实质被一些细枝末节所掩盖。



以下是编制统计表时必须注意的规则：

(1) 统计表的各种标题，特别是总标题的表达，应该十分简明、确切，概括地反映出表的基本内容。总标题还应该标明资料所属的地点和时间。

(2) 表中的主词各行和宾词各栏，一般应按先局部后整体的原则排列，即先列各个项目，后列总计。当没有必要列出所有项目时，可以先列总计，而后列出其中一部分的重要项目。

(3) 如果统计表的栏数较多，通常要加以编号。在主词和计量单位等栏，用(甲)(乙)(丙)等文字标明；宾词指标各栏，用(1)(2)(3)等数字编号。

(4) 表中数字应该填写整齐，对准位数。当数字为0或因数小可忽略不计时，要写上0；当缺乏某项资料时，用符号“…”表示；不应有数字时用符号“—”表示。

(5) 统计表中必须注明数字资料的计量单位。当全表只有一种计量单位时，可以把它写在表头的右上方。如果表中需要分别注明不同单位，横行的计量单位可以专设一栏；纵栏的计量单位，要与纵标目写在一起，用小字标写。

(6) 统计表的格式一般是“开口”式的，即表的左右两端不画纵线。

(7) 必要时，统计表应加注说明或注解。例如，某些指标有特殊的计算口径，某些资料只包括一部分地区，某些数字是由估算来插补等，都要加以说明。此外，还要注明统计资料的来源，以便查考。说明或注解一般写在表的下端。

### 任务卡 2

小青是某中学某班的班主任，为了鼓励学生好好学习，小青决定统计学生的成绩，制作从期中到期末考试成绩的条形统计图进行对比，对进步较大的学生进行物质奖励。假如你是小青，你会如何制作条形统计图？

---

---



## 知识卡 2 统计图

### 一、统计图

统计图是统计资料的另外一种表达方式，它可以简洁、直观地表示统计表中枯燥的数据，帮助我们众多的数据中发现规律，更迅速、更有效地传递信息，给人以明确而深刻的印象。

#### （一）统计图的结构

统计图基本包括以下几个部分：

- （1）标题。统计图一般包括图表标题、数值轴（X、Y）标题。
- （2）坐标轴和网格线。坐标轴和网格线构造了绘图区的骨架，借助坐标轴和网格线，我们可以更容易读懂统计图。
- （3）图表区和绘图区。统计表的所有内容都在图表区内，包括绘图区。统计图绘制在绘图区内。
- （4）图例。图例用来标明图表中的数据系列。当只有一个序列时，可以省略图例。有时统计图中的数据系列可能有多个，可以用不同颜色、形状的图例加以区别。

统计图的种类很多，常用的有用于辅助统计分析的直方图、趋势图和散点图，还有便于直观表现数据的柱形图、饼图、圆环图等。Excel 提供了 14 种标准图表类型，如图 3-13 所示，每种标准图表类型还可以包含几种不同的子类型，可以根据数据的类型以及自己的需要决定采用哪种图表来表达。



图 3-13 Excel 中的标准图表类型

## (二) 定类数据的统计图

适用于定类数据的统计图主要有饼图、条形图等。

### 1. 饼图

饼图也称圆形图，是用圆形及圆内扇形的面积来表示数值大小的图形。圆形图主要用于表示总体中各组成部分所占的比例，对于研究结构性问题十分有用。

在绘制饼图时，总体中各部分所占的百分比用圆内的各个扇形面积表示，这些扇形的圆心角度数是按各分组（即各类）的百分比乘  $360^{\circ}$  来确定的。例如，关注服务广告的人数占总人数的百分比为 25.5%，那么其对应的扇形圆心角度数就应为  $360^{\circ} \times 25.5\% = 91.8^{\circ}$ 。

### 微实训

#### 广 告 统 计

为研究广告市场的状况，一家广告公司在某城市随机抽取 200 人就广告问题做了邮寄问卷调查，其中的一个问题是“您比较关心下列哪一类广告？”

- A. 商品广告 B. 服务广告 C. 金融广告 D. 房地产广告 E. 招生招聘广告  
F. 其他广告

这里的变量就是“广告类别”，不同类型的广告就是变量值。调查数据经分类整理后形成频数分布，如表 3-9 所示。

表 3-9 某城市居民关注广告类型的频数分布

| 广告类型   | 人数 / 人 | 比 例   | 频 率   |
|--------|--------|-------|-------|
| 商品广告   | 112    | 0.560 | 56.0% |
| 服务广告   | 51     | 0.255 | 25.5% |
| 金融广告   | 9      | 0.045 | 4.5%  |
| 房地产广告  | 16     | 0.080 | 8.0%  |
| 招生招聘广告 | 10     | 0.050 | 5.0%  |
| 其他广告   | 2      | 0.010 | 1.0%  |
| 合计     | 200    | 1     | 100%  |

表 3-9 的频数分布可用饼图来表达，这可以借助 Excel 中的“图表”功能来操作（其实在 Microsoft Office 中的 Word、Excel 和 PowerPoint 等系统中均具有“图表”功能，这里仅就 Excel 做介绍）。具体步骤如下：



(1) 将表 3-9 中的数据录入 Excel 工作表中, 单击“插入”按钮, 然后单击“图表”, 出现图 3-14 所示对话框。



图 3-14 图表向导——图表类型

(2) 在对话框中选择“饼图”及“子图表类型(本例选择三维饼图)”, 然后单击“下一步”按钮, 出现图 3-15 所示对话框。



图 3-15 图表向导——图表源数据



(3) 在对话框中,“数据区域”一栏输入工作表中广告类型及频数所在的区域“\$A\$1:\$B\$6”,然后单击“下一步”按钮,出现图 3-16 所示对话框。

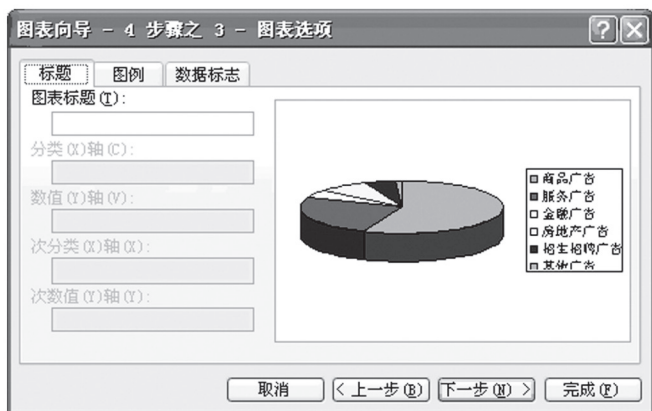


图 3-16 图表向导——图表选项

(4) 单击“下一步”按钮,出现图 3-17 所示对话框。然后选择图形输出的方式(本例选择“作为其中的对象插入”),单击“完成”按钮,即得所需的饼图(见图 3-18)。



图 3-17 图表向导——图表位置

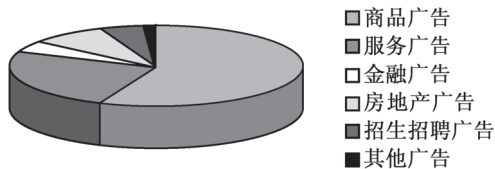


图 3-18 所需饼图

## 2. 条形图

条形图是用宽度相同的条形的高度或长短来表示频数分布的统计图形。

在表示定类数据的分布时,用条形图的高度来表示各类别数据的频数或频率。绘制时,各类别可以放在纵轴,称为条形图;也可以放在横轴,称为柱形图。根据表 3-9 绘制的条形图如图 3-19 所示。

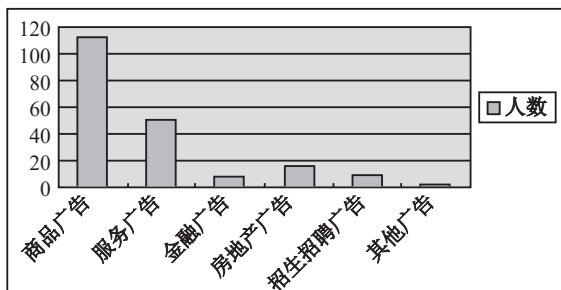


图 3-19 某城市居民关注广告类型频数分布的条形图

在 Excel 中，条形图的制作方法和饼图类似，这里不再赘述。

### （三）定序数据的整理与显示

由于定序数据中不仅有类别属性，还有顺序属性，因此，上述所介绍的饼图不能适用于定序数据。条形图虽然可以用来表达定序数据的频数分布，但要注意，其分组（即类别）必须按照顺序排列。

此外，条形图不仅可以表达定序数据的频数分布，而且可以表达累积频数分布，如图 3-20 和图 3-21 所示。

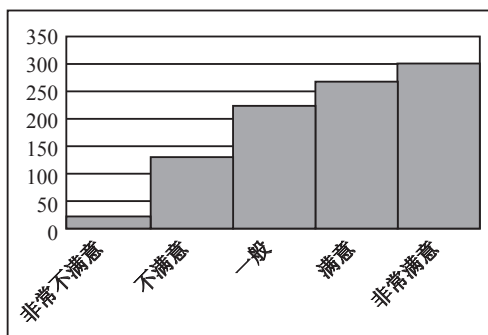


图 3-20 甲城市居民对甲城市住房状况评价的累积分布图

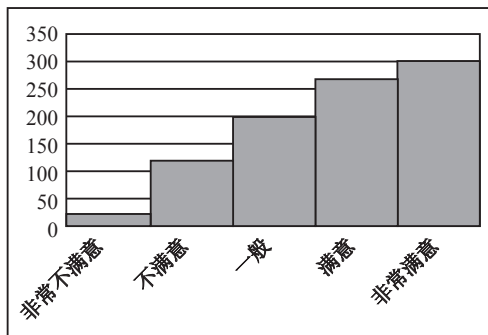


图 3-21 乙城市居民对乙城市住房状况评价的累积分布图

对于多总体定序数据，还可以采用环形图来表达。





环形图与圆形图类似，但又有区别。环形图中间有一个“空洞”，总体中的每一部分数据用环中的一段来表示。圆形图只能显示一个总体各部分所占的比例，而环形图则可以同时绘制多个总体的数据系列，每一个总体的数据系列为一个环，因此环形图可显示多个总体各部分所占的相应比例，从而有利于进行比较研究，如图 3-22 所示。

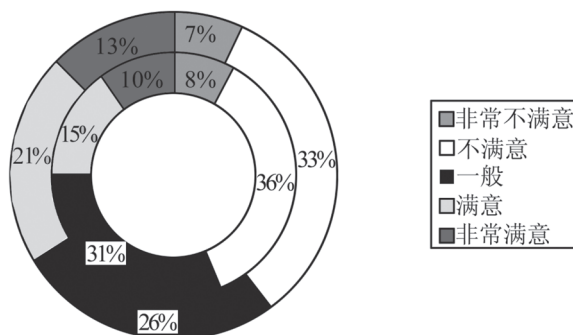


图 3-22 甲、乙两城市家庭对住房状况的评价

在图 3-22 中，外边的一个环表示乙城市家庭对住房状况各类评价所占的百分比，里边的一个环则为甲城市家庭对住房状况各类评价所占的百分比。

## 二、文字叙述

除了采用统计表和统计图作为统计整理结果的数据显示形式外，还可以采用文字叙述的形式。

文字叙述以文章、字句的形式叙述整理结果，这种方法主要用于口头汇报或发言。其目的是避免现场组织语言出现问题。在一般的统计分析报告，也多采用这种方式。



素质园地



## 职业基础知识测试

### 一、填空题

1. 统计数据整理就是根据\_\_\_\_\_, 对所收集到的\_\_\_\_\_进行科学的分类、汇总和加工, 使之\_\_\_\_\_的工作过程。统计数据整理既包括对\_\_\_\_\_所得到的原始资料进行整理, 也包括对加工过的\_\_\_\_\_, 即次级资料进行再整理。

2. 数据审核的内容包括\_\_\_\_\_。

3. 分组标志按照性质不同可分为\_\_\_\_\_, 按照\_\_\_\_\_可分为简单分组和复合分组。

4. 频数分布是指在\_\_\_\_\_的基础上, 把总体全部单位按组归类整理, 汇总出各组的总体单位数, 并将其按分组顺序加以排列而形成的\_\_\_\_\_。

5. 电子计算机对数据进行处理, 包括对数据的\_\_\_\_\_, \_\_\_\_\_、\_\_\_\_\_, 分类、排序、\_\_\_\_\_, \_\_\_\_\_、计算、传输、\_\_\_\_\_等工作。

### 二、简答题

1. 简述统计整理的意义。
2. 统计分组的作用有哪些?
3. 频数分布包括哪几种? 分别有哪些特征?
4. 简述统计汇总技术的分类。
5. 统计图的结构包括哪些?



## 职业核心能力训练

### 【实训目标】

- (1) 深入理解和熟练掌握频数分布表的制作、分组标志及类型。
- (2) 加强理论和实践的联系, 提高学生将统计方法应用于日常学习与生活的能力。

### 【具体要求】

请你对本班学生的统计学考试成绩进行统计整理, 分析本班学生该课程的考试情况。



(1) 根据本班实际人数情况确定样本容量。

(2) 将本班学生的考试成绩分为不及格、及格、中、良、优五组，编制一张频数分配表。

其中，学校规定：60 分以下为不及格，60~70 分为及格，70~80 分为中，80~90 分为良，90~100 分为优。

### 【实训记录】

### 【交流与评价】

| 项 目  | 成 绩                 |                    |                  |     |
|------|---------------------|--------------------|------------------|-----|
|      | 频数分布表的制作<br>( 30% ) | 样本容量适当性<br>( 30% ) | 结果合理性<br>( 40% ) | 总 分 |
| 自我评价 |                     |                    |                  |     |
| 同学评价 |                     |                    |                  |     |
| 教师评价 |                     |                    |                  |     |



## 项目四 掌握统计综合指标

### 知识目标 →

- 了解总量指标的含义和作用，理解总量指标的分类，了解总量指标的计算方法以及计算和应用总量指标应注意的问题
- 了解相对指标的意义和表现形式，理解计算和应用相对指标应注意的问题
- 掌握平均指标的概念和特点，了解平均指标的作用，理解权数及其作用
- 理解应用平均指标应注意的问题

### 能力目标 →

- 掌握总量指标的计算方法
- 准确运用相对指标进行计算
- 能够运用平均指标解决实际问题

### 思维导图 →

#### 掌握统计综合指标

##### 认识总量指标

总量指标的含义、作用和种类

总量指标的计算

##### 掌握相对指标

相对指标

应用相对指标应注意的问题

##### 掌握平均指标与标志变异指标

平均指标

标志变异指标



### 项目情境 →

小董下个月结婚，他想去商店购买两种糖混合成喜糖发给同事。商店里有巧克力糖、奶糖、酥糖、椰糖、玉米糖，每千克的价格分别为 30 元、18 元、15 元、12 元和 10 元。小董拿出预算的一半全部购买了巧克力糖。如果他希望他的喜糖包平均质量为 100 克 / 包，平均成本为 2 元 / 包，那么他应该用剩下的一半预算购买另外哪种糖？

通过上述描述，你认为应该运用什么统计综合指标来计算？请查找资料，说说你的看法。

---

---

---

---



## 任务一 认识总量指标

### 任务卡 1

某学校选出学生年龄、身高、体重、考试分数、生活费支出等数量标志，那么学生总体的标志总量为总年龄、总身高、总体重、考试总分数、生活费总支出等。你觉得上述标志总量恰当吗，为什么？

### 知识卡 1 总量指标的含义、作用和种类

#### 一、总量指标的含义

总量指标是反映社会经济现象在一定时间、地点条件下的总规模和总水平的指标。总量指标通常用绝对数的形式表示。总量指标的特点是：

- (1) 总量指标是绝对数形式，一定有计量单位。
- (2) 总量指标数值大小受总体范围大小的影响。总体范围大，总量指标数值则大；反之，总量指标数值就小。例如，一个省的 GDP 会大于该省某个市的 GDP 等。
- (3) 总量指标的计算只限于有限总体。

总量指标不仅可以表明现象的总体规模或水平，而且可以表明社会经济现象总体在不同的时间、地点条件下的数量增减变化的绝对量。

#### 二、总量指标的作用

总量指标是综合指标中最基本的指标，在统计分析和统计研究中具有十分



重要的作用。总量指标具体表现在以下几个方面：

(1) 总量指标是反映社会经济现象总体的基本状况和基本实力的统计指标。社会经济现象总体的基本情况和基本实力通常是使用总量指标反映出来的。大到一个国家、一个地区，小到一个企业、一个部门，其基本状况都需要总量指标来反映，如人口总量、资源总量、总产值和总投资等。

(2) 总量指标是计算相对指标和平均指标的基础。相对指标和平均指标一般是由两个总量指标对比得到的，所以它们是总量指标的派生指标。总量指标的设置及计算是否科学、正确，直接关系到相对指标和平均指标的准确性。

### 三、总量指标的种类

#### 1. 按照反映的内容分类

按其反映的内容不同，总量指标可分为总体单位总量指标和总体标志总量指标。

总体单位总量是表明总体单位数多少的总量指标，总体标志总量是表明总体单位某一数量标志值总和的总量指标。例如，对某单位全体职工的奖金情况进行研究，该企业“职工人数”是总体单位总量，全体职工的“奖金总额”是总体标志总量。

原始数据的总体单位总量和总体标志总量指标可借助于 Excel 中的“总和”和“计数”功能来汇总计算。这里要注意：

(1) 对于一个确定的总体而言，总体单位总量指标是唯一的，而总体标志总量指标则有许多。例如，某企业职工总体，其总体单位总量指标是“职工人数”，而“职工奖金总额”“职工工资总额”“总产量”等都是总体标志总量指标。

(2) 一个总量指标是总体单位总量还是总体标志总量，不是固定不变的，它随着研究目的和研究对象的不同而发生变化。例如，在研究某市中职学生的学习状况时，某市“中职学生人数”是总体单位总量指标，如果研究的目的是某市中职学校的基本情况，则各中职学校是总体单位，各中职学校的学生人数就是数量标志值，此时“全市中职学生人数”就是各总体单位标志值的和，自然属于总体标志总量指标。

#### 2. 按照反映的时间状态分类

按其反映的时间状态的不同，总量指标可分为时期指标和时点指标。





(1) 时期指标。时期指标是表明社会经济现象总体在一段时期内发展过程的总结果的总量指标。

(2) 时点指标。时点指标是表明社会经济现象总体在某一时刻(瞬间)数量状况的总量指标。

时期指标和时点指标的区别如下:

(1) 时期指标具有可加性,不同时期的指标数值相加表明较长时期的总量。时点指标不具有可加性,不同时点的指标数值相加没有实际意义。

(2) 时期指标的数值大小与时期长短有关,而时点指标的数值大小则与时间间隔长短没有直接关系。

(3) 时期指标的数值可以连续计数,而时点指标的数值只能间断计数。

### 3. 按照采用的计量单位分类

按其所采用的计量单位不同,总量指标可分为实物指标、价值指标和劳动量指标。

(1) 实物指标。实物指标是表明事物使用价值的总量指标。它采用的是实物计量单位,包括自然计量单位、度量衡单位和标准实物单位。

自然单位是按照被研究事物的自然属性来度量事物数量的计量单位。例如,机器以“台”为计量单位,企业以“个”为计量单位等。度量衡单位是按照统一的度量衡制度的规定来度量事物数量的一种计量单位。例如,布用“米”为计量单位,粮食用“千克”或“吨”为计量单位,发电量用“kW·h”为计量单位等。标准实物单位是按照统一折算的标准来度量被研究事物数量的一种计量单位。例如,把各种发热量不同的能源折合为 29 288 J/kg 的标准煤来计算,把各种度数不同的白酒按 65 度白酒为标准进行折算计量等。

(2) 价值指标。价值指标是表明事物价值量的总量指标,它以货币为单位进行计量。例如,工业总产值、商品销售额、国民生产总值、工资总额等都是以货币为单位计量的价值指标。价值指标能够使不能直接相加的产品或商品数量过渡到可进行加总,并综合说明不同使用价值的物品的总规模、总水平。该指标具有广泛的综合性和概括能力。

(3) 劳动量指标。劳动量指标是以劳动时间作为计量单位的总量指标。例如,“工时”“工日”“学时”等都属于劳动单位。企业管理中,劳动量指标常用



于编制企业的班组和个人生产计划，检查生产作业计划的完成情况，计算劳动消耗总量，核算产品成本及工人工资，计算劳动生产率等。

### 任务卡 2

请写出你所了解的总量指标的计算方法。

☐ 平衡关系推算法

☐ \_\_\_\_\_

☐ 插值估计法

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_

## 知识卡 2 总量指标的计算

### 一、总量指标的计算方法

总量指标数值都是通过对总体单位进行全面调查登记，采用直接计数、点数或测量等方法，逐步计算汇总得出的。总量指标的计算有直接计算法和间接推算法两种。

直接计算法是对研究对象用直接的计数、点数和测量等方法，登记各单位的具体数值加以汇总，得到总量指标。例如，统计报表或普查中的总量资料，基本上都是用直接计算法计算出来的。

间接推算法是利用社会经济现象之间的平衡关系、因果关系、比例关系或利用非全面调查资料推算总量指标的方法，如利用样本资料推断某种农产品的产量，利用平衡关系推算某种商品的库存量等。下面介绍几种经常使用的间接推算方法。

#### 1. 平衡关系推算法

平衡关系推算法是根据社会经济现象之间的平衡关系，利用已知总量指标来推算未知总量指标的方法。例如，在人口统计中，常住人口和现有人口之间通常存在如下平衡关系：

$$\text{常住人口} = \text{现有人口} + \text{暂时外出人口} - \text{暂时居住人口}$$



根据上述关系式可以对式中某一指标进行推算。

社会经济生活中广泛存在平衡关系。因而，平衡关系推算法是一种非常有用的方法。但应用平衡关系推算法时要求平衡关系中的各收支项目不能重复或遗漏，同时计算口径必须一致，否则容易出现错误。



想一想

## 2. 因素关系推算法

因素关系推算法是根据社会经济现象的因果关系，利用已知因素资料推算未知有关资料的方法。例如：

$$\text{工业总产值} = \text{工人人数} \times \text{工人劳动生产率}$$

$$\text{商品销售额} = \text{商品销售量} \times \text{价格}$$

这两个关系式中的某两项已知就可推算另一未知因素的数值。

## 3. 比例关系推算法

比例关系推算法是根据已知的某一时期、某一地区或某一单位的某种指标，推算另一时期、另一地区或另一单位的与其有着比例关系的相关指标，或者是根据总体组成部分的比例关系推算总体资料的方法。



想一想

## 4. 插值估计法

插值估计法是根据统计资料中若干已知项目的对应关系资料，来推算未知项目的对应关系的数值。常用的插值估计法有内插法和线性插值法。

(1) 内插法。内插法是根据若干已知资料来推算所缺数字资料。例如，利用平均发展速度估算逐年缺少的数字资料。其计算公式为

$$\bar{X} = \sqrt[n]{\frac{a_n}{a_0}}$$

式中， $\bar{X}$  为平均发展水平； $a_n$  为末期水平； $a_0$  为最初水平。

(2) 线性插值法。线性插值法是根据已知两项有关的对应关系资料估算第三项对应的未知数值。这种方法运用两点式原理，将已知的两项资料的对应关系表现为直线方程式：

$$\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0}$$

整理得



$$y = y_0 + \frac{y_1 - y_0}{x_1 - x_0} (x - x_0)$$

运用这一直线方程，即可根据已知资料求得未知数值。

与比例关系推算法不同，插值估计法是依据两项有关的对应关系资料估算第三项对应的未知数值，一般要比比例关系推算法按一项资料进行估算更接近实际。

## 二、运用总量指标应注意的问题

运用总量指标应注意以下问题：

(1) 正确统计和应用总量指标的一个首要问题，就是要对所统计的现象总体的范围和指标含义有一个明确的、科学的界限。

(2) 汇总时要注意现象的同类性。总量指标是说明总体的具体数值的，故有一定的计量单位，不同的计量单位有不同的特点。在计算实物总量指标时要特别注意现象的同类性，只有同类现象才能计算实物总量，不同类的现象不能计算实物总量。例如，钢与粮食的实物量就不能相加。同类现象的实物量应采用统一的计量单位，有些现象虽是同类，但实物内容差异很大，就需要区分其品种、观察其总量。

(3) 要有正确合理的计算方法。在统计总量指标时要根据研究目的，采取明确而合理的计算方法。

## 任务二 掌握相对指标

### 任务卡 1

小张的爸爸开了两家水果店，由于疫情原因，要关掉其中一家，这两家水果店的利润、占用资金、资金利用率都已知，小张的爸爸让小张来衡量关掉哪家。假如你是小张，你会选择两家的哪项指标进行比较，为什么？

---

---



## 知识卡 1 相 对 指 标

### 一、相对指标的含义和作用

#### 1. 相对指标的含义

相对指标又称相对数，是通过两个有联系的指标进行对比，以反映现象总体的数量结构、变化程度或现象之间的数量关系。例如，2020 年的 GDP 与 2019 年的 GDP 相比，其发展速度是 2.3%，这个指标数值就是通过两个时期的总量指标相比而得到的，用来反映国民经济的发展变化的速度。相对指标通过对比，把两个具体数值抽象化，使人们对事物有更清晰的认识。相对指标的计算公式为

$$\text{相对指标} = \frac{\text{对比数}}{\text{基数}}$$

#### 2. 相对指标的作用

相对指标是统计分析的重要方法，是反映现象之间数量关系的重要手段，在国民经济和社会发展问题的研究中具有很重要的作用。它主要表现在以下几个方面：

首先，相对指标可使原来不能直接相比较的指标进行对比。一些社会经济现象由于总体规模、结构和客观条件不同，其总量指标无法对比。例如，生产同一种类产品的两个企业，由于其规模大小不同，其工业总产值、职工总人数、利润总额、销售额和流动资金占用额等总量指标都是无法对比的。但通过计算有关的相对指标，如工业总产值增长速度、利润率和流动资金周转次数等，就可以对比两个企业的生产经营水平和经济效益的高低。

其次，相对指标是开展统计分析的重要工具。要分析一种社会经济现象，仅仅利用某一项指标，而不把有关指标联系起来进行比较分析，就难以把问题的实质和全貌反映出来。例如，工业产值的大小、计划执行的好坏、发展速度的快慢、各种比例关系协调与否只有通过比较才能反映出来。

最后，相对指标能够反映出现象之间相互联系的程度。例如，不同时间上的同类指标对比，计算动态相对数，可以研究现象的发展速度；不同地区、不



同单位的指标对比, 计算比较相对数, 可以研究现象发展的均衡程度; 同一总体内不同部分对比, 计算比例相对数, 可以反映现象之间的比例关系; 等等。

## 二、相对指标的种类和计算

根据研究目的的不同及对比基础的不同, 相对指标可分为结构相对指标、比例相对指标、比较相对指标、动态相对指标、强度相对指标和计划完成程度相对指标。

### 1. 结构相对指标

结构相对指标是利用统计分组法, 将总体划分为性质不同的部分, 然后用各部分的数值与总体数值对比得到相对数, 从而反映总体各组成部分占总体比重的大小。结构相对指标一般用百分数形式表示, 其计算公式为

$$\text{结构相对指标} = \frac{\text{总体中某一部分的数值}}{\text{总体的全部数值}} \times 100\%$$

结构相对指标的分子、分母指标, 可以是总体单位总量, 也可以是总体标志总量。一个总体中各部分的结构相对指标的和等于 100%。

结构相对指标的作用在于:

(1) 分析现象总体的内部构成状况, 说明事物的性质和特征。例如, 对某企业的全体职工按文化程度进行分组, 然后计算出不同文化程度的职工占全体职工的比重, 这样就可以了解该企业职工文化水平的高低, 为进一步提高职工的文化素质提供依据。在实际工作中, 往往需要计算出总体指标中一个分组指标占总体指标的比重, 这样的结构相对指标, 通常称为“率”, 如一级品率、工人出勤率等。

(2) 反映事物的变化过程和趋势。现象总体的变化经常是从结构变化开始的, 把不同时间上的结构相对指标进行对比分析, 可以说明现象的变化过程和趋势。

### 2. 比例相对指标

比例相对指标是用同一总体内部的两个不同组成部分之间的数值对比, 以反映各组成部分之间的数量关系, 用公式表示为

$$\text{比例相对指标} = \frac{\text{总体某一部分数值}}{\text{总体另一部分数值}}$$





比例相对指标用来反映同一总体内部的各组成部分之间的协调、均衡程度。例如，在人口总体中，男女之间的人口比例必须相互协调，可用“性别比例”指标来反映：

$$\text{性别比例} = \frac{\text{男性人口数}}{\text{女性人口数}}$$

例如，某地区2020年末人口总数为31 008万人，其中男性人口20 728万人，女性人口10 280万人，人口性别比例为201.6 : 100。

### 3. 比较相对指标

比较相对指标是用两个不同总体的同类指标数值对比，以反映某一现象在同一时间内不同空间条件下发展的均衡程度。其计算公式为

$$\text{比较相对指标} = \frac{\text{某一总体的某类指标数值}}{\text{另一总体的同类指标数值}}$$

比较相对指标也常用于不同地区、不同部门、不同单位之间的同类指标的对比，通过对比，有助于找出差距，揭露矛盾，挖掘潜力，改进工作，促进发展。

在计算比较相对指标时应注意，用于对比的两个指标所代表的现象必须是同质的、同类的。

### 4. 动态相对指标

动态相对指标是通过对某一指标在不同时间上的数值进行对比而得到的相对指标。它能说明同类事物在不同时间上的发展和变化程度，故也称为发展速度指标。其计算公式为

$$\text{动态相对指标} = \frac{\text{报告期指标数值}}{\text{基期指标数值}} \times 100\%$$

式中，报告期是指所要研究的时期；基期是指作为对比标准的时期。

动态相对指标对于分析社会经济现象的发展变化过程具有重要意义。

### 5. 强度相对指标

强度相对指标是通过对两个性质不同而又有密切联系的指标进行对比，以反映现象强度、密度或普及程度的相对指标。其计算公式为

$$\text{强度相对指标} = \frac{\text{某一总量指标的数值}}{\text{另一有联系而性质不同的总量指标的数值}}$$





强度相对指标多为有名数，用复合单位表示，即以对比的两个指标数值的计量单位组成在一起作为相对数的计量单位，如人口密度（人/平方千米）、人均 GDP（元/人）等。少数强度相对指标是无名数，如人口出生率（‰）、死亡率（‰）等。

值得注意的是，强度相对指标在表现形式上同平均指标十分相似，但它们有着实质性的差别，因为平均指标是总体标志总量除以总体单位总数，而强度相对指标是两个不同性质但又有密切联系的总体的指标之比。

## 6. 计划完成程度相对指标

计划完成程度相对指标是将现象在某一时期实际完成的数值与计划数值对比得到的相对数，一般用百分数表示。其计算公式为

$$\text{计划完成程度相对指标} = \frac{\text{实际完成数}}{\text{计划数}} \times 100\%$$

计算计划完成程度相对指标时，要求分子分母在指标含义、计算口径、计算方法、计量单位、时间长度和空间范围等方面应完全一致。

计划完成程度相对指标的特点如下：

（1）由于计划数总是衡量计划完成情况的标准，分子、分母不得互换。

（2）判断计划完成程度的好坏，要视指标的类型而定。对于正指标，如产量、产值、劳动生产率等，计划完成情况相对数大于 100% 才算超额；对于逆指标，如单位产品成本、流通费用率等，计划完成情况相对数小于 100% 才算超额；对于少数指标，如职工人数、工资总额、固定资产投资额等，是不允许突破计划的，这些指标的计划完成情况相对数以 100% 为好。

### 任务卡 2

小李的奶奶习惯每年体检一次，今年体检的结果有一项是全血低切相对指数偏低，这里的相对指数和我们统计学中的相对指数是一样的吗？说说你的看法。

---

---



## 知识卡 2 应用相对指标应注意的问题

相对指标是将社会经济现象的数字抽象化,因此,运用它来分析和说明问题时,应注意以下几个问题。

### 1. 必须遵循可比性原则

由于相对指标是由两个有联系的指标数值对比而得到的,因此,用来对比的两个数值是否可比,就成为计算相对数的前提条件。指标可比性的要求包括:指标所表明现象的经济内容是可比的,包括的范围相同,计算方法和计量单位也相同。指标的可比性问题是一个很复杂的问题。具体确定指标之间是否可比,要根据统计研究的任务,进行具体分析。既要防止过分强求可比性,而不敢对事物进行对比,又要注意防止过分忽视指标的可比性,盲目进行对比分析。有些指标在一种条件下是不可比的,但在另一种条件下又是可比的。

### 2. 相对指标应和绝对数结合运用

在运用指标时,只用绝对数就看不清事物的差异程度,只用相对指标则看不出事物差别的实际内容,只有将两者结合,才能使认识既深入又完整。

### 3. 选择好对比的基数

相对指标通过两个指标对比的结果来反映现象之间的数量关系,其中一个指标是对比标准,称为基数。在计算相对指标时,正确选择对比基数是进行对比分析的前提。如果对比的基数选择不当,将影响我们对现象的认识,有时甚至会产生错误的结论。在确定现象对比基数时,应从现象的性质和特点出发,选择那些能够反映事物内在联系和本质差异的指标作为对比标准。同时,确定对比基数还应考虑统计分析的任务和研究目的的要求,选择能够深刻说明问题的指标作为对比标准。

### 4. 各种相对指标要结合使用

一个相对指标只能说明现象某一方面的联系,但是社会经济现象不仅在静态上的联系是错综复杂的,而且与动态的发展也有相互联系。一个现象的变化往往影响着与之相联系的其他现象的变化。例如,产量结构变动会影响产值、原材料消耗和成本等一系列指标的变化。因此,只有把各种指标结合起来分析,



从不同的侧面看问题，才能做出全面正确的判断，避免认识的片面性。

## 任务三 掌握平均指标与标志变异指标

小明：平均指标、标志变异指标，这两个指标的名字真的有意思，而且平均指标还分成算术平均数、调和平均数、几何平均数、众数和中位数。为什么这样分呢？还真想打破砂锅问到底。等把这几个“数”弄明白后，可以到爸爸那儿炫耀一番了。

### 任务卡 1

你了解平均指标吗？说说平均指标和平均数的区别。

### 知识卡 1 平均指标

#### 一、平均指标的概念和作用

##### 1. 平均指标的概念

平均指标是反映统计数据（即总体单位标志）一般水平的统计指标，又称统计平均数。其特点是将各统计数据的差异抽象化，代表了全部统计数据的一般水平，反映了现象总体的综合数量特征。

##### 2. 平均指标的作用

（1）反映了总体分布的集中趋势。统计数据的分布是分散的，但一般会集中在平均指标的附近。也就是说，越靠近平均指标，统计数据越多，而两端的数据较少。这说明总体分布是从两边向中间集中，中心是平均数，因此平均数反映了一组数据向某一中心值靠拢的趋势，故可以直接把平均指标称为集中



趋势。

(2) 利用平均指标便于进行对比分析。平均指标作为一个代表值,不仅使个别单位的离差相互抵消,而且不受总体单位多少的影响,因此,便于用来对社会经济现象总体进行对比分析。一方面,平均指标可以用来在不同部门、不同地区和不同单位之间进行对比;另一方面,平均指标可以反映某现象的水平在不同时期的变化,以便说明这些现象的发展趋势或规律性。

(3) 利用平均指标可以分析现象之间的依存关系。在对现象进行分组的基础上,结合应用平均指标,可以观察现象之间存在的依存关系。例如,分析劳动生产率水平与平均工资水平的关系,分析工业企业固定资产与平均产值的关系等。

(4) 平均指标是制定定额的依据。因为平均指标反映总体各单位的一般水平,表现现象总体单位的综合特征。它是先进水平与落后水平综合作用的结果,所以平均指标是制定各种定额的依据。

(5) 利用平均指标可以做数量上的推算。在统计推算中,常利用部分单位的平均数推算总体平均数,根据总体某个标志的平均数与总体单位数可以推算总体标志总量。

平均指标按其计算方法的不同,可分为算术平均数、调和平均数、几何平均数、众数和中位数等。

## 二、算术平均数

算术平均数又称均值,是总体标志总量除以总体单位总量所得的平均数。算术平均数是计算平均指标最常用的方法和最基本的形式,适用于刻画定距数据的一般水平。其计算公式为

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

在实际工作中,由于所掌握的统计资料不同,利用上述公式计算时,又分为简单算术平均数和加权算术平均数两种。

(1) 简单算术平均数。简单算术平均数适用于未分组的定距数据。设有  $n$  个数据  $x_1, x_2, \dots, x_n$ , 则这  $n$  个数据的算术平均数为



$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$



想一想

(2) 加权算术平均数。加权算术平均数指根据经过分组所得的总体分组资料计算的算术平均数, 即各组标志值或组中值乘以相应的各组单位数求出各组标志总量, 并加总求得总体标志总量, 然后除以各组单位数之和所得的算术平均数, 其计算公式为

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_n f_n}{f_1 + f_2 + f_3 + \cdots + f_n} = \frac{\sum x_i f_i}{\sum f_i}$$

式中,  $x_i$  为各组标志值或组中值;  $f_i$  为各组单位数。

其中各组的单位数即频数, 又称为权数, 故利用上式计算的平均数称为加权算术平均数。

由于算术平均数具有以上性质, 在应用时具有以下特征:

- (1) 可用于同类现象一般水平的比较。
- (2) 对于一个数列来讲, 算术平均数只有一个。

(3) 在组距数列中, 特别是具有开口组的数列, 计算算术平均数时, 利用组中值作为该组的代表值, 存在假定性, 其结果只是一个接近实际平均数的近似值。

(4) 算术平均数受极端值的影响。在数列中, 如果有一个数特别大或特别小时, 计算出来的平均数会降低它的代表性。

### 三、调和平均数

在计算平均数时, 由于缺乏总体单位的资料, 不能直接采用算术平均数的方法计算平均数, 应采用调和平均数。调和平均数是各个变量倒数的算术平均数的倒数。

调和平均数是平均指标的一种, 与算术平均数一样, 因给定资料的形式不同, 也分为简单调和平均数与加权调和平均数。

#### 1. 简单调和平均数

简单调和平均数的计算公式为



$$H = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}$$

式中,  $H$  为调和平均数;  $x$  为各单位标志值;  $n$  为标志值的项数。

【例 4-1】某厂生产甲、乙、丙 3 种不同的设备, 单位成本分别为 0.125 万元/台、0.2 万元/台和 0.5 万元/台。

(1) 若 3 种设备各生产 1 台, 试计算 3 台设备的平均单位成本。

(2) 假设报告期 3 种设备各发生成本费用 1 万元, 平均每台设备的成本又是多少?

解: (1) 应按简单算术平均数计算:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{0.125 + 0.2 + 0.5}{3} = 0.275 \text{ (万元/台)}$$

(2) 这时, 作为计算平均指标基础的总体单位数不是 3 台, 而是 15 台, 即

$$\frac{1}{0.125} + \frac{1}{0.2} + \frac{1}{0.5} = 15 \text{ (台)}$$

故要计算这 15 台的平均单位成本, 就必须采取简单调和平均数计算:

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{1+1+1}{\frac{1}{0.125} + \frac{1}{0.2} + \frac{1}{0.5}} = 0.2 \text{ (万元/台)}$$

## 2. 加权调和平均数

简单调和平均数是在频数分布中各标志值所对应的标志总量均为一个单位的情况下, 求平均数时应用。但在大多数情况下, 分配数列中各标志值所对应的标志总量往往不相等, 在这种情况下, 要求平均数, 就必须应用加权调和平均数。

加权调和平均数的计算公式为

$$H = \frac{m_1 + m_2 + \cdots + m_n}{\frac{m_1}{x_1} + \frac{m_2}{x_2} + \cdots + \frac{m_n}{x_n}} = \frac{\sum m_i}{\sum \frac{m_i}{x_i}}$$



式中,  $m_i$  为各变量值;  $x_i$  为所对应的权数。

【例 4-2】某年甲乙两个工厂某种产品的销售情况如表 4-1 所示。

表 4-1 某年甲乙两个工厂某种产品的销售情况

| 工厂名称 | 销售额 / 元 | 销售价格 / (元 · 件 <sup>-1</sup> ) |
|------|---------|-------------------------------|
| 甲    | 5 000   | 1                             |
| 乙    | 4 800   | 1.2                           |

试计算两个工厂产品的平均价格。

解: 产品的平均价格是产品销售额与销售量之比, 但是, 在没有掌握商品销售量资料的情况下, 要求计算某种产品的平均价格, 就不能直接用算术平均数的方法计算, 而必须先求出产品销售量, 再以产品销售量除产品销售额。因此, 产品的平均价格应按下式计算:

$$\text{产品的平均价格} = \frac{5\,000 + 4\,800}{\frac{5\,000}{1.0} + \frac{4\,800}{1.2}} \approx 1.09 \text{ (元)}$$

调和平均数的特点如下:

- (1) 调和平均数也易受极端值的影响。
- (2) 当组距数列有开口组时, 所遇到的问题与计算算术平均数是一样的。

## 四、几何平均数

几何平均数是  $n$  个变量值乘积的  $n$  次方根, 它常用于计算平均比率或平均速度。凡是变量值的连乘积等于总比率或总速度的现象都适合用几何平均数来计算其平均比率或平均速度。几何平均数的计算公式为

$$G = \sqrt[n]{x_1 x_2 \cdots x_n} = \sqrt[n]{\prod x_i}$$

式中,  $G$  为几何平均数;  $x_i$  为各个变量值;  $n$  为变量个数;  $\prod$  为连乘号。

在变量数值比较多的情况下, 运算较为复杂, 可借助计算器或采用对数计算。若采用对数计算, 则

$$\lg G = \frac{1}{n} (\lg x_1 + \lg x_2 + \cdots + \lg x_n) = \frac{\sum \lg x_i}{n}$$



想一想

几何平均数的对数等于变量值对数的算术平均数, 这里也可以看出几何平





均数和算术平均数之间的关系。

## 五、众数

众数是指在某一总体中出现次数最多的标志值,或者在变量数列中具有最多次数的变量值,用  $M_0$  表示。例如,某生产小组的 8 个工人日产量分别为 17 件、18 件、16 件、16 件、15 件、16 件、16 件、17 件,其中 16 件出现的次数最多,所以把 16 称为众数。

如果各个变量值都不相同,那就没有众数。

如果项数很多,则可以编制组距次数分布数列,求众数的近似值。

众数的概念通俗易懂,不必进行很复杂的计算,同时不受个别极大值或极小值的影响,所以在统计分析中可用于以下情况:

(1) 通过其频数的多少来反映研究总体频数分配的集中状况。其频数在总体单位总量中所占比重越大,表明被研究的总体集中程度越大,众数对总体的代表性也越大。

(2) 在没有必要或不可能计算平均数和中位数时,可利用众数说明问题。

(3) 众数与算术平均数结合,可使分析更全面。

众数的确定方法有两种:一种是在单项数列的情况下确定,另一种是在组距数列的情况下确定。

在单项数列中,确定众数比较简单,只需找出次数最多的标志值即可。

在组距数列中,众数不能直接看出,要通过公式计算,然后确定众数 ( $M_0$ ) 的近似值,其计算公式为

$$M_0 = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times d \quad (\text{下限公式})$$

$$M_0 = U - \frac{\Delta_1}{\Delta_1 + \Delta_2} \times d \quad (\text{上限公式})$$

式中,  $L$  为众数组的下限;  $U$  为众数组的上限;  $\Delta_1$  为众数组次数与前一组次数之差;  $\Delta_2$  为众数组次数与后一组次数之差;  $d$  为众数组的组距。

【例 4-3】某企业 500 名工人按工资分组,所得组距分布如表 4-2 所示,试计算工人工资的众数。



表 4-2 某企业工人工资分布

| 工人按工资分组 | 工人数 / 人 |
|---------|---------|
| 40~50 元 | 45      |
| 50~60 元 | 80      |
| 60~70 元 | 240     |
| 70~80 元 | 90      |
| 80~90 元 | 45      |
| 合计      | 500     |

解：首先需确定众数组。表中 60~70 元这一组的人数最多，故该组为众数组。将表 4-2 中的数据代入下限公式计算：

$$M_0 = 60 + \frac{240 - 80}{(240 - 80) + (240 - 90)} \times 10 = 65.16 \text{ (元)}$$

或代入上限公式计算：

$$M_0 = 70 - \frac{240 - 90}{(240 - 80) + (240 - 90)} \times 10 = 64.84 \text{ (元)}$$

众数的计算需满足一个条件，即单位数比较多，而且有明显的集中趋势。如果单位数很少或单位数多但无明显的集中趋势，是无法计算众数的。

## 六、中位数

### 1. 中位数的概念及特点

中位数是指将统计数据按大小顺序排列后，居于中间位置的数据。中位数的概念表明，在数列中，应有一半的数据小于中位数，一半的数据大于中位数。

中位数不受极值的影响，通过中位数，可以从一个侧面反映频数分配的集中状况。在被研究的总体中，若大部分总体单位标志值比较集中，只有少数分散在一端，且标志值极大或极小，若用平均数表示一般水平，易受极端变量的影响，这时，用中位数就比用平均数更能确切地反映频数分配的集中趋势。

### 2. 中位数的计算方法

(1) 由未分组资料确定中位数。在数据未分组的情况下，计算中位数的方法是先把各数据按大小顺序排列，如果项数是奇数，则按  $(n+1)/2$  的公式计算即得中位数的位置；如果项数是偶数，直接取中间两个数的平均数。



【例 4-4】两个公司的职工文化程度资料如表 4-3 所示，试比较两个公司的职工文化程度。

表 4-3 两个公司的职工文化程度分布

| 文化程度 | 人 数   |       | 向上累计频数 |       |
|------|-------|-------|--------|-------|
|      | 甲 公 司 | 乙 公 司 | 甲 公 司  | 乙 公 司 |
| 高中   | 5     | 2     | 5      | 2     |
| 专科   | 50    | 28    | 55     | 30    |
| 本科   | 42    | 64    | 97     | 94    |
| 硕士   | 15    | 25    | 112    | 119   |
| 博士   | 3     | 3     | 115    | 122   |
| 合计   | 115   | 122   |        |       |

解：甲公司的职工总人数为 115 人，按文化程度高低排序后，中位数位置为  $(115+1) \div 2=58$ ，根据向上累计频数可知中位数在第三组，故甲公司的职工文化程度中位数为“本科”。同理，乙公司职工文化程度中位数也为“本科”。这说明两个公司职工的文化程度的一般水平是相同的。

(2) 在组距分布情况下确定中位数。在组距分布情况下中位数的计算比较复杂，要运用公式计算，计算公式为

$$M_e = L + \frac{\frac{\sum f}{2} - S_{m-1}}{f_m} \times i \quad (\text{下限公式})$$

$$M_e = U - \frac{\sum f - S_{m+1}}{f_m} \times i \quad (\text{上限公式})$$

式中， $M_e$  为中位数； $L$  为中位数所在组的下限； $f_m$  为中位数所在组的次数； $\sum f$  为总次数； $i$  为中位数所在组的组距； $U$  为中位数所在组的上限； $S_{m-1}$  为中位数所在组下限以前各组的累计次数； $S_{m+1}$  为中位数所在组上限以后各组的累计次数。

## 七、运用平均指标需要注意的问题

在计算和分析平均数的过程中，为了使平均数能正确地表明客观事物的性



质和规律，在应用平均指标时要注意以下几个问题。

### 1. 计算平均指标必须以同质总体为基础

平均指标是同质总体中各单位某一数量标志值在一定时间和空间的条件下所达到的一般水平的综合指标。平均指标是将现象总体各单位某一数量标志差异抽象化，反映现象总体在一定时空条件下所达到的一般水平或代表水平。如果各个单位在类型上是异质的，特别是对社会生产关系来说存在根本区别，那么利用平均数就会抹杀现象之间的本质区别。因此，计算平均指标必须以同质总体为基础。

### 2. 进行平均分析时必须结合组平均数和次数分配

根据同质总体计算的平均数虽然是科学的，并可以正确地反映客观实际情况，但只能说明总体的一般水平，往往不能深入地说明现象之间变化的原因。因为总体单位在某一标志上是同质的，但在其他标志上存在差别，为了更深入地说明、认识事物的全面情况，必须利用分组法，将总体各单位分组，把组平均数和总体平均数结合起来进行分析，用组平均数来补充总体平均数，这样才能深入地揭示所研究对象发展变化的原因。同时，由于平均数是把一个数列标志值的差异抽象化后得到的一般水平的数值，不能反映出数列的分布情况。通过结合变量数列分析就可以看出各单位标志值与一般水平的差别及其分布特征。

### 3. 以变异指标补充说明平均指标

平均指标将总体各单位标志值的差异抽象化，只能反映这些标志值的一般水平，所以平均指标只能反映各单位某一数量标志的共性，而不能反映它们之间的差异性。仅用平均指标还不能全面描述总体标志值分布的特征，而标志变异指标则能弥补这个不足。它综合反映总体各单位标志值的差异，从另一方面说明总体的特征。所以，应该以变异指标补充说明平均指标，这样才能较全面地反映出总体的数量特征。



### 任务卡 2

王叔家里有两个孩子，王叔把两个孩子的期末考试成绩列入一个表格（见表 4-4）进行分析。

表 4-4 大宝和二宝的期末考试成绩

| 姓 名 | 成绩 / 分 |    |    |    |    |    |
|-----|--------|----|----|----|----|----|
|     | 语文     | 数学 | 物理 | 化学 | 政治 | 英语 |
| 大宝  | 95     | 90 | 65 | 70 | 75 | 85 |
| 二宝  | 110    | 70 | 95 | 50 | 80 | 75 |

通过计算，两个孩子的平均成绩为 80 分，然而王叔说二宝偏科严重并把二宝教育了一番。你知道王叔教育二宝的原因吗？说说你的看法。

## 知识卡 2 标志变异指标

### 一、标志变异指标的含义和作用

#### 1. 标志变异指标的含义

平均指标代表了统计数据的一般水平，它消除了各个数据之间的差异。而标志变异指标则表明总体各个数据之间的差异程度，或者说是离散程度，所以又称为标志变动度。它与平均指标的作用是相辅相成的。

#### 2. 标志变异指标的作用

（1）平均指标是衡量平均指标代表性大小的依据。平均指标作为全部统计数据的代表值，其代表性的高低取决于数据之间的差异程度，因此，它是与标志变异指标直接相关的，这种关系表现为：数据间的标志变异指标越大，平均指标的代表性越小；反之，标志变异指标越小，平均指标的代表性越大。

例如，有两组学生，每组各有 4 个学生，每个学生某门课程的期末考试成绩



绩分别为：

甲组（分）：70，75，82，85

乙组（分）：60，64，85，93

甲、乙两组的平均成绩都是 78 分，但两组学生的分数的离散程度不同，甲组相对小一些，因而甲组平均成绩的代表性就高于乙组。标志变异指标的这一作用有助于我们分析判断现象发生的均衡性或稳定程度。

（2）平均指标是抽样调查和相关分析中需要使用的一个重要指标。它是影响和计量抽样误差的一个重要因素，也是计算相关关系的一个重要依据。

## 二、标志变异指标的计算方法

标志变异指标的计算方法一般有四种：全距、平均差、标准差和标准差系数。

### 1. 全距

全距是指总体中单位标志值的最大值与最小值的差距，即数列中两个极端数值之差，故又称极差。全距用公式可表示为

$$\text{全距}(R) = \text{最大标志值} - \text{最小标志值}$$

一般来说，全距越小说明标志变动值越集中，全距越大说明标志变动值越分散。如前例中甲、乙两组学生的全距分别为

$$R_{\text{甲}} = 85 - 70 = 15 \text{ (分)}$$

$$R_{\text{乙}} = 93 - 60 = 33 \text{ (分)}$$

计算结果表明，乙组学生平均成绩的代表性小于甲组学生平均成绩的代表性。

全距的计算较为简单，它只考虑到两头最大和最小的数值，所以易受极端值的影响。它没有考虑数列中间几个变量值的变动，因而不能全面反映标志变动的程度。

如果统计资料经过整理并形成组距分配数列，则全距的近似值为

$$\text{全距}(R) = \text{最高组的上限} - \text{最低组的下限}$$

### 2. 平均差

平均差是指总体中各数据与平均数离差的绝对值的算术平均数。在统计中把每个统计数据与平均数的差（ $x_i - \bar{x}$ ）称为离差，离差的平均数就是平均差。



由于离差有正有负，离差之和等于零，即  $\sum(x_i - \bar{x}) = 0$ ，因此，计算离差的平均数即平均差，可以采用离差的绝对值，即  $|x_i - \bar{x}|$  相加之和来计算。平均差一般用  $A \cdot D$  表示。

计算平均差的公式有两种：

(1) 简单平均式。对于未分组数据可采用简单平均式来计算平均差，即

$$A \cdot D = \frac{\sum |x_i - \bar{x}|}{n}$$

式中， $n$  为总体单位数即离差的项数； $x_i$  为统计数据； $\bar{x}$  为平均数。

【例 4-5】表 4-5 是某车间两个生产小组的日产量资料，试通过平均差比较两组平均数的代表性。

表 4-5 某车间两个生产小组的日产量资料

| 日产量 $x_i$ / 件 | 第 一 组                         |                            | 日产量 $x_i$ / 件 | 第 二 组                         |                            |
|---------------|-------------------------------|----------------------------|---------------|-------------------------------|----------------------------|
|               | 标志值与平均数的离差<br>$x_i - \bar{x}$ | 离差绝对值<br>$ x_i - \bar{x} $ |               | 标志值与平均数的离差<br>$x_i - \bar{x}$ | 离差绝对值<br>$ x_i - \bar{x} $ |
| 20            | -50                           | 50                         | 67            | -3                            | 3                          |
| 40            | -30                           | 30                         | 68            | -2                            | 2                          |
| 60            | -10                           | 10                         | 69            | -1                            | 1                          |
| 70            | 0                             | 0                          | 70            | 0                             | 0                          |
| 80            | 10                            | 10                         | 71            | 1                             | 1                          |
| 100           | 30                            | 30                         | 72            | 2                             | 2                          |
| 120           | 50                            | 50                         | 73            | 3                             | 3                          |
| 合计            | 0                             | 180                        | 合计            | 0                             | 12                         |

解：两个小组工人的平均日产量都为  $\bar{x} = 70$  件。

根据简单平均式计算出两组工人日产量的平均差：

$$\text{第一组的平均差 } A \cdot D_1 = \frac{\sum |x_i - \bar{x}|}{n} = \frac{180}{7} \approx 25.7 \text{ (件)}$$

$$\text{第二组的平均差 } A \cdot D_2 = \frac{\sum |x_i - \bar{x}|}{n} = \frac{12}{7} \approx 1.7 \text{ (件)}$$





计算结果表明, 第一组的平均差大于第二组, 因此第一组的平均数的代表性比第二组小。

(2) 加权平均式。若数据已经整理成频数分布, 则可采用加权平均式计算平均差, 即

$$A \cdot D = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i}$$

式中,  $f_i$  为各个组的频数, 是计算平均差的权数。

平均差不同于全距, 它是根据所有变量值计算的, 因此能综合反映总体中各单位标志值的离散程度。平均差越大说明标志变动度越大, 平均数的代表性越小; 反之, 平均差越小说明标志变动度越小, 平均数的代表性越大。

### 3. 标准差

标准差也称为均方根差, 是指各数据与其平均数离差的平方的算术平均数的平方根。它是测定标志变动度最主要的指标, 用希腊字母  $\sigma$  表示。

标准差的意义与平均差类似, 也是数据与平均指标的平均离差, 所不同的是平均差平均的是离差绝对值, 而标准差平均的是离差平方。标准差彻底解决了正负离差不能相加的问题, 它是就各组离差平方求平均数后再开平方而求得的。标准差可以准确地、综合地反映总体的离散程度, 同时, 从算术平均数的性质来看, 各标志值与其算术平均数的离差平方和为最小。由于标准差具有如此之优点, 在实际工作中, 一般都用它来测定总体离散程度。

标准差的计算公式有两种:

(1) 简单平均式。如果数据未经整理分组, 可采用简单平均式来计算标准差。其计算公式为

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

式中,  $\sum (x_i - \bar{x})^2$  为离差平方和;  $n$  为总体单位数即离差项数。

这里把  $\sum (x_i - \bar{x})^2 / n$  称为方差, 用  $\sigma^2$  表示。

(2) 加权平均式。如果数据被整理成了频数分布, 就要采用加权平均式来计算标准差。其计算公式为



$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$$

式中,  $f_i$  为次数即计算平均离差的权数。

标准差越大说明标志变动的程度越大, 因而平均数的代表性越小; 反之, 标准差越小说明标志变动的程度越小, 因而平均数的代表性越大。

(3) 是非数据标准差。是非数据是指统计数据只表现为“是”或“非”两种情况。例如, 某人口总体中每个人的性别、某班每个同学的成绩(按“及格”“不及格”记)、某厂每件产品的质量(按“合格”“不合格”记)等都属于是非数据。也就是说, 是非数据是指每个数据或者具有某种属性(记为“是”), 或者不具有某种属性(记为“非”)。

由于是非数据属于定性数据, 不能用数字表示, 因而也不能据以计算其平均数和标准差, 所以有必要将其虚拟数量化。一般将反映“是”的数据记为 1, 将反映“非”的数据记为 0, 同时, 用  $n$  表示数据总数, 用  $n_1$  表示总体中“是”的数据个数, 用  $n_0$  表示总体中“非”的数据个数, 用  $p$  表示总体中“是”的数据所占比重, 用  $q$  表示总体中“非”的数据所占比重, 则

$$p = \frac{n_1}{n}$$

$$q = \frac{n_0}{n}$$

由于  $n = n_1 + n_0$ , 故  $p + q = 1$ 。

是非数据的频数分布及平均数计算如表 4-6 所示。

表 4-6 是非数据的分布及平均数计算

| 标志值 $x_i$ | 单位数 $f_i$ | $x_i f_i$ | $\frac{f_i}{\sum f_i}$ | $x_i \cdot \frac{f_i}{\sum f_i}$ |
|-----------|-----------|-----------|------------------------|----------------------------------|
| 1         | $n_1$     | $n_1$     | $p$                    | $p$                              |
| 0         | $n_0$     | 0         | $q$                    | 0                                |
| 合计        | $n$       | $n_1$     | 1                      | $p$                              |



则

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{n_1}{n} = p$$

可见，是非数据的平均数就是“是”的数据所占的比例。

是非数据的方差为

$$\begin{aligned}\sigma^2 &= \frac{\sum (x - \bar{x})^2 f}{\sum f} = \frac{(1-p)^2 n_1 + (0-p)^2 n_0}{n} \\ &= \frac{q^2 n_1 + p^2 n_0}{n} = \frac{q^2 n_1}{n} + \frac{p^2 n_0}{n} \\ &= q^2 p + p^2 q = pq(q+p) \\ &= pq = p(1-p)\end{aligned}$$

是非数据的标准差为

$$\sigma = \sqrt{pq} = \sqrt{p(1-p)}$$

由此可见，是非数据的标准差就是  $p$  和  $q$  乘积的平方根。

例如，某批产品有 1 000 件，其中合格品 900 件，则产品合格率为 90%。将“合格”记为“是”或“1”，“不合格”记为“非”或“0”，则该批产品质量状况的平均数为

$$\bar{x} = 90\%$$

方差和标准差分别为

$$\sigma^2 = 90\% \times 10\% = 9\%$$

$$\sigma = \sqrt{9\%} = 30\%$$

#### 4. 标准差系数

标准差系数是指标准差与算术平均数的比值，通常用百分数表示。其计算公式为

$$V_\sigma = \frac{\sigma}{\bar{x}} \times 100\%$$

标准差虽能正确反映标志变动度的大小，但利用它来比较平均数的代表性是有限的。只有在平均数相等的情况下，才能直接进行比较。如果平均数不相



等,相差很大,就不能直接进行比较。因为标准差数值的大小不仅受各单位标志值之间变异程度的影响,而且受标志值平均水平高低的影响。为了比较不同水平的平均数的代表性,就必须以平均数去除标准差,以消除标准差受平均水平的影响。



素质园地



## 职业基础知识测试

### 一、填空题

1. 总量指标是反映现象总体在一定时间、地点条件下的\_\_\_\_\_的指标。总量指标通常用\_\_\_\_\_的形式表示。
2. 2020 年底,小明的爸爸在银行的储蓄存款数是\_\_\_\_\_指标。
3. 时期指标是表明社会经济现象总体\_\_\_\_\_的总量指标。时点指标是表明社会经济现象总体\_\_\_\_\_的总量指标。
4. 根据研究目的的不同及对比基础不同,相对指标可分为\_\_\_\_\_,\_\_\_\_\_,比较相对指标、\_\_\_\_\_,\_\_\_\_\_。
5. 算术平均数又称\_\_\_\_\_,是总体标志总量除以\_\_\_\_\_所得的平均数。算术平均数是计算平均指标最常用的方法和最基本的形式,适用于刻画\_\_\_\_\_的一般水平。

### 二、简答题

1. 简述总量指标的分类。
2. 什么是相对指标?它有什么作用?
3. 什么是平均指标?它分为哪几种?
4. 什么是标志变异指标?它有什么作用?
5. 运用平均指标时需要注意哪些问题?



## 职业核心能力训练

### 【实训目标】

- (1) 深入理解和熟练掌握平均指标的原理和方法。
- (2) 加强理论和实践的联系,提高将统计方法应用于具体生产的能力。



### 【具体要求】

- (1) 对某食品企业三条生产线的日生产量情况进行调查。
- (2) 选取该食品企业三条生产线为调查对象，分析研究各生产线的日生产能力及标准差，找出最能代表该产品日生产能力的生产线。
- (3) 根据各生产线实际生产情况搜集所需数据，样本容量不小于 20。

### 【实训成绩评定】

实训的成绩要结合抽样设计的合理性和该产品的实际生产情况综合考虑。

### 【实训记录】

### 【交流与评价】

| 项 目  | 成 绩                |                     |                     |     |
|------|--------------------|---------------------|---------------------|-----|
|      | 标准差的准确性<br>( 30% ) | 搜集数据的全面性<br>( 30% ) | 调查结果的合理性<br>( 40% ) | 总 分 |
| 自我评价 |                    |                     |                     |     |
| 同学评价 |                    |                     |                     |     |
| 教师评价 |                    |                     |                     |     |



## 项目五 掌握抽样技术

### 知识目标 →

- 掌握抽样调查的含义和特点，了解抽样调查的作用
- 理解抽样调查中常用的若干基本概念
- 掌握简单随机抽样条件下的抽样误差、抽样平均误差、抽样极限误差的概念，掌握抽样误差的影响因素，了解抽样估计的置信度
- 理解参数估计的含义
- 理解影响样本容量的因素

### 能力目标 →

- 能够运用抽样调查方法进行统计
- 能够运用概率抽样的方法进行统计

### 思维导图 →

#### 掌握抽样技术

##### 认识抽样调查

抽样调查概述

抽样调查中的几个基本概念

##### 掌握抽样误差

抽样误差的含义和影响因素

抽样平均误差

##### 理解参数估计

参数估计的理论基础和总体参数估计

样本容量的确定



### 项目情境 →

某市教育局接到上级任务，要求全面了解全市小学生的学习情况。该市有 300 所小学，共有 240 000 名学生；这些小学分布在全市的 5 个行政区中，其中重点小学有 30 所，一般小学有 240 所，较差的小学有 30 所。现在要从全市小学生中抽取 1 200 名学生进行调查，以了解全市小学生的学习情况。假如你是该市教育局局长，你将采用什么方法抽样调查？

---

---

---

---





## 任务一 认识抽样调查

### 任务卡 1

有句名言是：“你不必吃完整个鸡蛋，才知道它是坏的。”在这句话中，有没有用到抽样调查的知识呢？说说你的看法。

### 知识卡 1 抽样调查概述

#### 一、抽样调查的含义和特点

##### 1. 抽样调查的含义

抽样调查是非全面调查，它是按随机原则从调查对象（即总体）中抽取部分单位作为样本进行调查，用调查所得指标数值对总体的指标数值做出具有一定可靠性的估计和推断的一种统计调查方法。随机原则是指在抽样时排除主观上有意识地抽取调查单位，每个受试单位以概率均等的原则，随机地分配到实验组与对照组，使每一个单位都有一定的机会被抽中。例如，从一定面积的小麦中，通过随机抽样，抽取若干块地实割实测，计算平均亩产，以此来推断全部面积的小麦亩产。还可以利用亩产乘以全部面积推算出全部面积的小麦总产量。再如，对一批产品进行质量检测时，从全部产品中抽取部分产品进行检测，计算合格率，以此来推断全部产品的合格率。还可以利用全部产品数量乘以合格率，推算出全部合格产品数量。可见，抽样调查不仅是一种科学的收集资料的方法，而且是一种科学的估计和推断的方法。



## 2. 抽样调查的特点

(1) 以部分推断总体。抽样调查是一种非全面调查,它作为一种进一步推断的手段,目的仍在于认识总体的数量特征。抽样调查资料如果不进行抽样推断,这种资料就不会有什么价值。这里存在手段与目的、局部与整体的矛盾。这种矛盾在现实生活中大量存在。例如,检测几克棉花纤维的强度,能否判断整批棉花纤维的强度;对几粒种子进行催芽试验,能否判断该品种整批种子的发芽率;等等。如果在方法上不能解决这类问题,那么统计的认识活动就要受到限制,统计科学也很难得到发展。抽样推断原理解决了这一矛盾,它科学地论证了样本指标与相应的总体参数之间存在内在的联系,两者的误差分布也是有规律可循的,并提供了一套利用抽样调查的部分信息来推断总体数量特征的方法,大大提高了统计分析的认识能力,为信息的采集和开发开辟了一条崭新的道路。

(2) 按随机原则抽取调查单位。这是抽样调查与其他非全面调查(如重点调查、典型调查)的主要区别之一。重点调查和典型调查的调查单位是由调查者有意识选取的,抽样调查的调查单位选取则不受调查者主观意志的影响。抽样调查为什么要遵循随机原则呢?首先,遵循随机原则才能使调查对象中的每个单位有同等机会被抽中,当抽取足够多的单位时,才能使被抽中单位的次数分布类型与调查对象相同,从而增强被抽中单位对总体的代表性。其次,遵循随机原则才能计算抽样误差,从而达到推断总体的目的。

(3) 抽样推断是运用概率估计的方法。利用样本指标来估计总体参数,在数学上运用概率估计法,而不是运用确定的数学分析法。因为样本数据和总体参数之间并不存在严格对应的自变量和因变量的关系,它不能利用一定的函数关系来推算总体参数。那么用这样的样本指标值代表相应的总体指标值,其可靠程度究竟有多大,这就是概率估计所要解决的问题。例如,我们不知道全校学生的平均身高是多少,现在抽取若干学生作为样本,并计算样本的平均身高为 168 厘米,又求得以 168 厘米代表全校学生平均身高的误差不超过 1.5 厘米的概率保证程度不低于 95%。如果这一估计的可靠程度被认为已经满足分析工作的要求,我们就可以用 168 厘米作为全校学生身高的平均水平,否则就要改善抽样组织,重新进行抽样调查,以提高结论的可靠程度,这是概率估计的基本思路。



(4) 抽样调查的误差可以事先计算并加以控制。以样本指标估计相应的总体指标肯定存在一定误差,但它与其他统计估算不同,其抽样误差范围可以事先通过有关资料计算出来,并且可以采取必要的措施,如改善样本、扩充样本容量、采用更适当的抽样组织形式等来控制这个误差范围,还可以保证抽样调查结果达到所要求的可靠程度。这一特点体现了抽样调查方法的科学性,这些都是其他估算方法所办不到的。

## 二、抽样调查的方法

由于其所具有的特点和优势,抽样调查在社会经济调查中被越来越广泛地应用。抽样调查主要有两种方法:非概率抽样和概率抽样。需要根据不同情况选用不同的方法。

### 1. 非概率抽样

非概率抽样是用主观(非随机的)方法从总体中抽选单元进行调查,它是一种快速、简便且费用低的抽选样本的方法。但是非概率抽样具有很大的风险,因为主观选定的样本很难肯定其对总体是否具有代表性。所以一般情况下非概率抽样不用来对总体进行推断,而更多的是为了达到以下目的:一是用来形成一种想法,二是为设计开发概率抽样调查方案做准备工作(如初步确定出某一总体方差等),三是在后续步骤中帮助理解概率抽样调查结果。

非概率抽样主要有随意抽样、志愿者抽样、判断抽样和配额抽样。

(1) 随意抽样。单元的抽选以无目的、随意的方式进行,几乎没有或完全没有计划。随意抽样假定总体是同质的,如果总体单元都相似,那么可以抽选任何单元入样。“街道拦截”访问法是随意抽样的典型例子,它是指访问员就某个问题抽选一个碰巧路过的人进行调查。但是这样的调查结果受两个方面的影响:一是谁在抽样,不同的人抽样选择哪个过路人是有倾向性的;二是在抽样时谁碰巧路过,就同一个问题而言,中职学校的老师和一名商人的理解往往是大相径庭的。

(2) 志愿者抽样。在使用这种方法时,被调查者都是志愿者。例如,2020年我国遭受新冠肺炎疫情后,为研究新冠肺炎疫苗,需要进行人体接种试验,但这种试验不可能采取随机抽样,而只能采取志愿者抽样的方法。媒体座谈讨论节目通常也从志愿者中抽选参与者。



(3) 判断抽样。在采用这种方法时, 应选择一些很了解总体的专家来决定总体中作为样本的单元。这些专家以过去对总体构成或行为的了解, 有目的地挑选他认为“有代表性”的样本。这种方法在探索性研究中很有用。

(4) 配额抽样。这是最常见的一种非概率抽样。例如, 一个总体有 100 名男性和 100 名女性, 要从中抽取 20 个人的样本, 且男、女各 10 人。这种方法是从各个子总体中选取特定数量的单元(配额), 而对于如何完成这些配额并没有规定。配额抽样比其他非概率抽样稍好一些, 因为它强制性地要求样本包含不同子总体的单元。

## 2. 概率抽样

概率抽样在抽取样本时不带有任何倾向性, 它通过从总体中随机抽选单元来避免这种偏差, 因而对总体的推断更具代表性。但是, 随机不意味着随意, 特别是访问员不能随意选择被调查者, 否则样本将受个人意愿的影响。概率抽样有两条基本准则: 第一, 单元是随机抽取的; 第二, 调查总体中的每个单元都有一个非零的入样概率。

概率抽样方法主要有简单随机抽样、系统随机抽样、分层随机抽样、整群随机抽样、多阶随机抽样和比例随机抽样, 常用的为前四种。

(1) 简单随机抽样。简单随机抽样是一种一步抽样法, 它要求在调查总体中不加任何分组、划类、排队等, 完全随机抽取  $n$  个调查单位作为样本。在简单随机抽样中, 总体中的每个单位都有相同的被抽中的概率, 这个概率记作

$$p = \frac{n}{N}$$

(2) 系统随机抽样。系统随机抽样也称为机械随机抽样或等距随机抽样。它是先将总体中各单位按一定的标志排队, 然后每隔一定的距离抽取一个单位构成样本。例如, 某小区居住着 4 000 户居民, 为调查居民的食盐摄入量, 需抽取 40 户居民进行调查, 平均每 100 户抽取 1 户。为此, 调查机构到当地派出所, 根据户籍资料将居民户从 1~4 000 编好号码, 先在 1~100 号中随机抽取 1 个号码, 再将这个号码加上 100 的整数倍, 共 40 户构成样本。例如, 在第一组中抽取了 5 号, 于是 5, 105, 205, ..., 3 805, 3 905 共 40 户被抽取出来作为样本。

(3) 分层随机抽样。分层随机抽样又称为类型随机抽样、分类随机抽样。它是按照某一标志先将总体分成若干组(类), 其中每一组(类)称为一层, 再



在层内按简单随机抽样方法进行抽样。例如,某地区有平原、丘陵和山地三种地形,粮食产量统计中需要采用抽样调查的方式进行。由于三种地形地貌的粮食单产差异较大,因此若采用简单随机抽样,可能使调查结果与实际情况相差甚远。例如,抽中的多是平原地区,有可能调查数值远远高于实际产量;而抽中的多是山地,有可能调查数值远远低于实际产量。为消除上述可能的偏差,先将该地区分为平原、丘陵、山地三个组,然后在每个组内按简单随机抽样的方法抽取调查地块,构成样本。

(4) 整群随机抽样。整群随机抽样是先将总体按某一标志分成若干组,其中每个组称为一个群,以群为单位进行简单随机抽样,然后对抽到的群内的每个单位都进行调查。例如,在了解某地区职工的家庭生活状况时,按居民委员会分群,一个居民委员会为一群,对抽中的居民委员会所辖每户职工家庭一一进行调查。

(5) 多阶随机抽样。多阶随机抽样是用两个或多个连续的阶段抽取样本的过程。第一阶段抽取的单元称为初级或一级抽样单元,第二阶段抽取的单元称为次级或二级抽样单元,以此类推。每个阶段抽取的单元在结构上是不同的。常见的多阶抽样是二阶抽样。例如,第一阶段抽样抽出小的地理小区,第二阶段再从被抽中的小区内抽选住所,然后进行调查。

(6) 比例随机抽样。与大小成比例的比例随机抽样是一种使用辅助信息,从而使入样概率不相等的抽样技术。如果一个总体中每个单位的大小或规模变化很大,而且这些大小是已知的,就可以在抽样中采用这些信息,以提高统计效率。例如,对农场进行某方面的调查,因为农场很多,而且规模大小相差很大,所以在调查中就要使规模大的农场入样概率要大一些,而规模较小的农场的入样概率要小一些。

### 三、抽样调查的作用

由于抽样技术具有费用低、时效强、准确度高、应用范围广等优点,抽样技术广泛应用于各个领域。

(1) 有些现象是无法进行全面调查的,为了测算全面资料,必须采用抽样调查的方法。例如,对无限总体不能进行全面调查;又如,有些产品的质量检查具有破坏性,如轮胎里程试验、灯管耐用时数检验等,这些调查所用的测试





手段对产品具有破坏性，不可能进行全面调查，只能采用抽样调查。

(2) 采用抽样调查既可以节省人力、物力、费用和时间，提高调查结果的实效性，又能达到和全面调查同样的目的和效果。从理论上讲，有些现象可以进行全面调查，但实际上没有必要或很难办到，这时就要采用抽样调查的方式。例如，一家物业公司计划购买一批灯泡，需要调查灯泡的使用寿命。从理论上讲可以做到对每一个灯泡进行全面调查，但是调查的范围太大，数量太多，实际上难以做到，也没必要，可以从中随机抽取样本，进行破坏性试验。

(3) 抽样调查的结果可以对全面调查的结果进行检查和修正。全面调查涉及面广、工作量大、参加人员多、调查结果容易出现差错。因此，在全面调查（如人口普查、工业普查）之后进行抽样复查，根据复查结果计算差错率，并以此为依据检查和修正全面调查结果，从而提高全面调查的质量。

(4) 抽样调查可用于工业生产过程的质量控制。在工业产品成批或大量连续生产的过程中，利用抽样调查可以检验生产过程是否正常，及时提供信息，进行质量控制，保证生产质量稳定。

(5) 利用抽样调查原理，可以对某些总体的假设进行检验，以判断这种假设的真伪，从而决定行动的取舍。例如，某地区上年粮食平均亩产 400 千克，本年抽样调查结果表明，粮食平均亩产 408 千克，这是否意味着粮食生产水平提高了呢？我们还不能下这个结论，最好通过假设检验，检验这两年粮食平均亩产是否存在显著性统计差异，然后才能判断该地区今年粮食亩产是否高于上年水平。

总之，抽样调查是一种科学实用的调查方法，它不仅广泛应用于自然科学领域，也越来越多地应用于社会经济现象数量方面的研究。随着抽样理论的发展，抽样技术进步和完善以及广大统计工作者业务水平的提高，抽样调查在社会经济统计中的应用将更加普及。

### 任务卡 2

为了了解初一（4）班 39 名同学的视力情况，我们从中抽取 10 名同学进行检查。这属于哪种抽样调查方法？说说你的看法。

---

---



## 知识卡 2 抽样调查中的几个基本概念

在学习抽样推断方法的过程中，我们常会遇到一些名词、术语。为了掌握抽样调查推断的方法，首先要明确抽样调查推断过程中常用的几个基本概念。

### 一、总体和样本

#### 1. 总体

总体也称全及总体，是指所要认识的研究对象的全体。总体单位数通常是很大的，甚至是无限的，因此有必要组织抽样调查。一般用大写字母  $N$  来表示总体单位数。

#### 2. 样本

样本又称抽样总体或子样，它是从总体中随机抽取出来的，作为代表总体的那部分单位组成的集合体。样本的单位数总是有限的，一般用小写字母  $n$  来表示。与总体的单位数  $N$  相比， $n$  是很小的数。

总体总是唯一确定的，但作为观察对象的样本就不是这样。从一个总体可以抽取很多个样本，每次可能抽到哪个样本不是确定的，也是不唯一的，而是可变的。

例如，某城市有 30 万户居民，我们采用抽样调查方法，从全部住户中抽取 3‰ 即 900 户进行调查，则全部住户构成总体， $N=300\ 000$ ，被抽中的 900 户构成样本， $n=900$ 。总体 30 万户是唯一的，但哪 900 户构成样本呢？答案有很多种，即样本不是确定的，也不是唯一的。

### 二、总体指标和样本指标

#### （一）总体指标

总体指标又称总体参数，它是反映总体数量特征的指标。当确认了总体后，总体参数就是一个确定的数值，是一个常数，但它往往又是未知的，需要通过抽样推断加以估计。分析一个总体往往需要用到多个总体指标，它们从不同角度反映了总体分布的基本状况和主要特征，常见的总体指标有总体平均数  $\bar{X}$ 、总体标准差  $\sigma(X)$  和总体方差  $\sigma^2(X)$ 、总体成数  $P$  等。





### 1. 总体平均数

总体平均数是代表了总体单位一般水平的统计指标, 是根据总体各个标志值计算而得到的平均数。其计算公式如下。

(1) 总体未分组。

$$\bar{X} = \frac{\sum X}{N}$$

式中,  $\sum X$  为总体标志总量;  $N$  为总体单位总数。

(2) 总体已分组。

$$\bar{X} = \frac{\sum XF}{\sum X}$$

式中,  $\sum XF$  为总体各组变量值。

### 2. 总体标准差和总体方差

总体标准差是根据总体各单位标志值计算的标准差, 它反映了总体数据对总体平均数的偏离程度。总体标准差用  $\sigma$  表示, 它的平方就是总体方差, 用  $\sigma^2$  表示。计算公式如下。

(1) 总体未分组。

$$\sigma^2 = \frac{\sum (X - \bar{X})^2}{N}, \quad \sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

(2) 总体已分组。

$$\sigma^2 = \frac{\sum (X - \bar{X})^2 F}{\sum F}, \quad \sigma = \sqrt{\frac{\sum (X - \bar{X})^2 F}{\sum F}}$$

### 3. 总体成数

在社会经济统计中, 把某种社会经济现象的全部总体单位分为具有某一标志的单位和不具有某一标志的单位两组, 总体成数是指总体中具有某种标志表现或不具有某种标志表现的单位数占总体单位数的比重。在总体中,  $P$  表示总体中具有某种性质的单位数在总体单位数中所占的比重,  $Q$  表示总体中不具有某种性质的单位数在总体单位数中所占的比重。

在总体  $N$  个单位中, 有  $N_1$  个单位具有某种性质,  $N_0$  个单位不具有某种性



质,  $N = N_1 + N_0$ , 则计算公式为

$$P = \frac{N_1}{N}, \quad Q = \frac{N_0}{N} = \frac{N - N_1}{N} = 1 - P$$

#### 4. 总体是非标志

是非标志指的是如果总体中的标志表现只有“是”和“非”两种(例如, 产品质量的标志表现为合格和不合格, 人口性别的标志表现为男性和女性, 可以“是”的标志表现表示为1, 而“非”的标志表现表示为0), 那么成数  $P$  就可以视为  $(0, 1)$  分布的相对数, 并可以计算相应的平均数和方差(或标准差), 其计算公式为

$$\begin{aligned}\bar{X}_P &= \frac{\sum XF}{\sum X} = \frac{0 \times N_0 + 1 \times N_1}{N_0 + N_1} = \frac{N_1}{N} = P \\ \sigma_P^2 &= \frac{(0-P)^2 N_0 + (1-P)^2 N_1}{N_0 + N_1} = \frac{P^2 N_0 + Q^2 N_1}{N} \\ &= P^2 Q + Q^2 P = PQ(P+Q) = PQ = P(1-P)\end{aligned}$$

即总体为是非标志或者  $(0, 1)$  分布的时候, 平均数就等于它本身  $P$ , 方差就等于  $P(1-P)$ 。

### (二) 样本指标

在抽样推断中, 总体指标的意义和计算方法是明确的, 但往往总体指标的具体数值事先是未知的, 需要用样本指标来估计。样本指标又称样本统计量, 是根据样本数据计算出来的用来估计总体指标的综合指标。常见的样本指标有样本平均数  $\bar{X}$ 、样本标准差  $s$  和样本方差  $s^2$  及样本成数  $p$  等。

#### 1. 样本平均数

样本平均数又称抽样平均数, 它是样本中各单位标志值的平均数, 一般用  $\bar{X}$  表示。其计算公式如下。

(1) 样本未分组。

$$\bar{x} = \frac{\sum x}{n}$$

式中,  $\sum x$  为样本标志总量;  $n$  为样本单位总数。



(2) 样本已分组。

$$\bar{x} = \frac{\sum xf}{\sum f}$$

式中,  $\sum xf$  为样本标志总量;  $\sum f$  为样本单位总数。

## 2. 样本标准差和样本方差

样本标准差就是根据样本中各单位标志值计算的标准差, 它反映了样本数据对样本平均数  $\bar{x}$  的偏离程度。样本标准差用  $s$  表示, 它的平方是样本方差, 用  $s^2$  表示。其计算公式如下。

(1) 样本未分组。

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}, \quad s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

(2) 样本已分组。

$$s^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}, \quad s = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}}$$

## 3. 样本成数

样本成数也称抽样成数, 它是指样本中具有某种标志表现或不具有某种标志表现的单位数占样本单位数的比重。在总体中, 设  $p$  表示样本中具有某种性质的单位数在样本单位数中所占的比重,  $q$  表示样本中不具有某种性质的单位数在样本单位数中所占的比重。在样本  $n$  个单位中, 有  $n_1$  个单位具有某种性质,  $n_0$  个单位不具有某种性质,  $n = n_1 + n_0$ , 则计算公式为

$$p = \frac{n_1}{n}, \quad q = \frac{n_0}{n} = \frac{n - n_1}{n} = 1 - p$$

## 4. 样本是非标志

样本是非标志指的是如果样本中的品质表现只有“是”和“非”两种, 样本为是非标志或者 (0, 1) 分布的时候, 可以计算相应的平均数和方差 (或标准差)。其计算公式如下。

样本是非标志的平均数:



想一想

$$\bar{x}_p = \frac{n_1}{n} = p$$

样本是非标志的方差：

$$s_p^2 = p(1-p)$$

总体指标和样本指标计算公式的比较如表 5-1 所示。

表 5-1 总体指标和样本指标计算公式的比较

| 指标类型 | 平均数                                                               | 方差                                                                                               | 成数                  | 成数方差     |
|------|-------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|---------------------|----------|
| 总体指标 | $\bar{X} = \frac{\sum X}{N}$ 或 $\bar{X} = \frac{\sum XF}{\sum F}$ | $\sigma^2 = \frac{\sum (X - \bar{X})^2}{N}$ 或 $\sigma^2 = \frac{\sum (X - \bar{X})^2 F}{\sum F}$ | $P = \frac{N_1}{N}$ | $P(1-P)$ |
| 样本指标 | $\bar{x} = \frac{\sum x}{n}$ 或 $\bar{x} = \frac{\sum xf}{\sum f}$ | $s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$ 或 $s^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}$         | $p = \frac{n_1}{n}$ | $p(1-p)$ |

任务二 掌握抽样误差

任务卡 1

老师上新课前总是先抽查几个同学对以前课程的掌握情况，每次抽查时小亮总是提心吊胆。请你说说小亮被抽到的概率有多大？和哪些因素有关？



## 知识卡 1 抽样误差的含义和影响因素

### 一、抽样误差的含义

抽样误差是指调查所获得的统计数据与调查总体未知真实数据之间的差别。它包括登记性误差和代表性误差两种。

登记性误差是在调查过程中由于主客观原因引起的登记性差错所造成的误差。例如，由于测量、记录、计算造成的误差，由于被调查者所报材料不实或调查者有意虚报、瞒报资料造成的误差，等等。这类误差可以通过搞好调查的宣传组织工作，提高调查人员素质，采用电子计算机技术等措施加以解决。

代表性误差是用样本指标数值去推算总体指标数值时，由于样本各单位的结构情况不足以代表总体特征所产生的误差。这类误差又可分为两种。一种是由于没有遵守随机原则而造成的误差，通常称为系统性误差。例如，调查者有意挑选较好或较差的单位做样本。另一种误差是即使遵守了随机原则，也会由于被抽取的样本各种各样，导致样本内部各单位的分布比例结构与总体实际分布状况有偶然性的差异，从而使不同的随机样本得出不同的估计量，造成样本指标数值与总体指标数值之间产生差距，如抽样平均数与总体平均数的离差（ $\bar{x} - \bar{X}$ ），抽样成数与总体成数的离差（ $p - P$ ）等。这类误差通常称为抽样误差或随机误差。可见，只要遵循随机原则，系统性误差是可以控制在一定范围内的。同样，用样本估算总体，两者之间总是要出现差距的，所以抽样误差是抽样法本身所固有的误差，它不可避免，也无法消除，但是可以控制。

### 二、抽样误差的影响因素

抽样误差的大小受以下几个因素的影响。

#### 1. 样本单位数

在其他条件不变的情况下，抽样单位数越多就越能反映总体，误差就越小。当抽样单位数接近总体单位数时，抽样调查已接近全面调查，抽样误差接近零；反之，抽样单位数越少，抽样误差越大。

#### 2. 总体被研究标志变异程度

在其他条件不变的情况下，所研究总体的标志变异程度越小，说明总体各



单位标志值之间的差异越小，这样抽样指标与总体指标之间的误差也越小。相反，若总体被研究标志变异程度大，则抽样指标与总体指标之间的误差也大。

### 3. 抽样组织方式

采取不同的抽样组织方式，所抽出的样本对于总体的代表性也不相同，因此抽样组织方式影响抽样误差的大小。在实践中，我们可以利用不同抽样组织方式下抽样误差的大小来判断不同方式的有效性。

### 4. 抽样方法

抽样方法有重复抽样和不重复抽样两种。在其他条件相同时，不重复抽样的抽样误差一般小于重复抽样的误差，这是因为不重复抽样避免了总体单位的重复中选，因而能更准确地反映总体结构。

#### 任务卡 2

你了解抽样误差吗？请上网查找资料，说说抽样误差不可以避免的原因，并举一个抽样误差的例子。

## 知识卡 2 抽样平均误差

### 一、抽样平均误差的含义

在一定组织方式下进行一次抽样调查，根据统计研究的目的和任务，可以取一个样本，也可以取多个样本。在抽取多个样本时，就其中每个样本来说，都有其相应的样本指标。

由于样本是按随机原则抽取的，故在同一总体中，按相同的抽样数目，可以抽出许多相同和不同的样本，而每次抽出的样本都可以计算出相应的抽样平均数、抽样成数和抽样误差。即从理论上说可以计算出很多个抽样误差，它们带有偶然性，有的可能是正误差，有的可能是负误差；有的可能大一些，有的可能小一些。为了用样本指标推算总体指标，我们需要计算这些抽样误差的平



均数，这就是抽样平均误差。抽样平均误差用以反映抽样误差的一般水平。抽样平均误差越大，表示样本的代表性越低；抽样平均误差越小，表示样本的代表性越高。抽样平均误差说明了样本指标与总体指标的相差范围，是推算总体指标的依据。

## 二、抽样平均误差的计算方法

抽样平均误差反映样本指标和总体指标的平均离差程度，不但能说明样本指标的代表性大小，也是以样本指标估计和推断总体指标时计算极限误差的基础。

对于简单随机抽样来说，抽样平均误差的定义公式等价于如下计算公式。

重复抽样条件下，有

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \quad (5-1)$$

$$\mu_p = \sqrt{\frac{P(1-P)}{n}} \quad (5-2)$$

不重复抽样条件下，有

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)} \quad (5-3)$$

$$\mu_p = \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n}{N}\right)} \quad (5-4)$$

在运用公式(5-1)~公式(5-4)时，需要注意以下问题。

(1) 公式中的标准差 $\sigma$ 和成数 $P$ 是总体的标准差和成数，而总体指标是未知的，一般可以用样本的标准差 $s$ 和成数 $p$ 或用历史资料来代替。

(2) 在不重复抽样条件下，如果抽样单位数相对较少，而总体单位数相对较多，则 $\left(1 - \frac{n}{N}\right)$ 趋近于1，即不重复抽样的抽样平均误差和重复抽样的抽样平均误差近似相等。在实际工作中，常用不重复抽样的方法来抽取调查单位，进行总体指标的推断，而用重复抽样的方法来计算抽样平均误差。



想一想



### 三、抽样极限误差

抽样平均误差说明了某一总体的所有样本指标与总体指标间误差的平均数,但在实际工作中往往只能抽取一个样本,因此实际抽样误差一般不会等于抽样平均误差,可能大于或小于抽样平均误差。所以,用抽样平均误差无法准确地推断总体指标值。此时,就需要引入抽样极限误差。

抽样极限误差是指样本指标与总体指标之间抽样误差的最大可能范围,又称为允许误差。以  $\Delta_{\bar{x}}$  和  $\Delta_p$  分别表示样本平均数和样本成数的抽样极限误差,则有

$$|\bar{x} - \bar{X}| \leq \Delta_{\bar{x}}$$

$$|p - P| \leq \Delta_p$$

实际上,抽样极限误差是一个可能而非完全肯定的范围,这个可能范围的大小是与可能性大小相对应的。在抽样推断中,表示这个可能性大小的概念称为置信度,习惯上称为可靠程度、把握程度或概率保证程度,用  $F(Z)$  表示,其中的  $Z$  值称为概率度。

抽样理论已经证明:样本平均数  $\bar{x}$  服从以总体平均数  $\bar{X}$  为中心的正态分布,该正态分布的标准差就是抽样平均误差  $\mu_{\bar{x}}$ 。因此,根据正态分布中变量取值区间与概率的关系可知:样本平均数落在  $(\bar{X} \pm \mu_{\bar{x}})$  范围内的可能性为 68.27%;落在  $(\bar{X} \pm 2\mu_{\bar{x}})$  范围内的可能性为 95.45%;落在  $(\bar{X} \pm 3\mu_{\bar{x}})$  范围内的可能性为 99.73%。具体如图 5-1 所示。

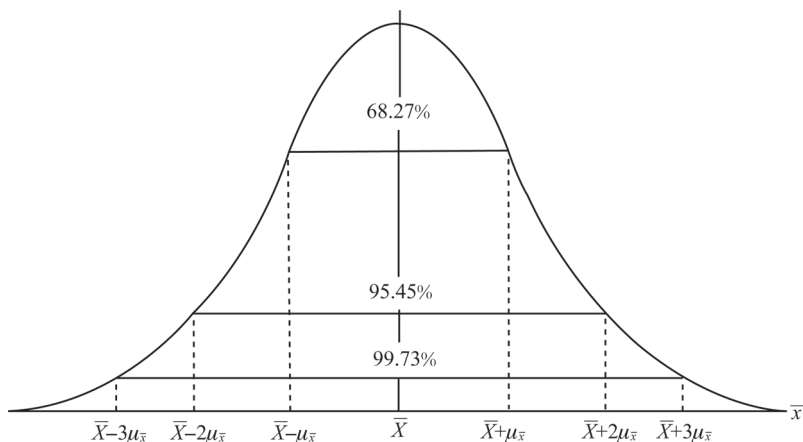


图 5-1 样本平均数的分布



上述结论等价于：在 68.27% 的置信度下，样本平均数的抽样极限误差等于其抽样平均误差；在 95.45% 的置信度下，样本平均数的抽样极限误差等于其抽样平均误差的 2 倍；在 99.73% 的置信度下，样本平均数的抽样极限误差等于其抽样平均误差的 3 倍。由此可见，样本平均数的抽样极限误差可以用其抽样平均误差的倍数来衡量，具体计算公式如下。

$$\Delta_{\bar{x}} = Z\mu_{\bar{x}}$$

同理，样本成数的抽样极限误差如下。

$$\Delta_p = Z\mu_p$$

上述两个公式中， $Z$  值是由抽样推断时给定的置信度所决定的，其对应关系可查询标准正态分布表。实际工作中常用的几个  $Z$  值和置信度如表 5-2 所示。

表 5-2 常用的几个  $Z$  值和置信度

| $Z$ 值 | 置 信 度  |
|-------|--------|
| 1     | 68.27% |
| 1.64  | 89.90% |
| 1.96  | 95.00% |
| 2     | 95.45% |
| 3     | 99.73% |
| 5     | 99.99% |

从表 5-2 中可以看出，置信度越大， $Z$  值就越大，抽样极限误差也越大，抽样推断的精确度越低。所以，在抽样推断中，要求达到 100% 的置信度是不可能的。但另一方面，置信度小了，推断的可靠性也小，又会影响推断本身的价值。因此，在进行抽样推断时，应将置信度要求与推断的精确度要求结合起来考虑。



## 任务三 理解参数估计

### 任务卡 1

周末小张去超市买了一些水果，他让爸爸估算自己花了多少钱，并且告诉爸爸花的钱数在 100 元和 150 元之间。请说明估算钱数是不是区间估计，为什么？

### 知识卡 1 参数估计的理论基础和总体参数估计

#### 一、参数估计的理论基础

抽样估计是建立在概率论的大数定律基础上的，大数定律为参数估计提供了数学依据。

大数定律是关于大量的随机现象具有稳定性质的法则。它指出如果被研究的总体由大量的相互独立的随机现象所组成，而且每个因素对总体的影响都相对较小，那么对这些大量因素加以综合平均，因素的个别影响将相互抵消，从而显现出它们共同作用的倾向，使总体具有稳定的性质。

联系到抽样推断来看，大数定律说明：如果随机变量总体存在有限的平均数和方差，则对于充分大的抽样单位数  $n$ ，可以以几乎趋近 1 的概率来期望抽样平均数与总体平均数的绝对离差为任意小。设  $\varepsilon$  为任意小的正数，则

$$\lim_{n \rightarrow +\infty} p(|\bar{x} - \bar{X}| \leq \varepsilon) = 1$$

式中， $\bar{x}$  为抽样平均数； $\bar{X}$  为总体平均数； $n$  为样本单位数。

这就从理论上揭示了样本和总体之间的内在联系，说明随着抽样单位数  $n$



的增加, 抽样平均数  $\bar{x}$  有接近总体平均数  $\bar{X}$  的趋势。

大数定律说明了抽样平均数趋近于总体平均数的趋势, 为抽样推断提供了重要的依据。但是大数定律未能解释抽样平均数与总体平均数的离差不超过一定范围的概率大小问题, 这一问题由中心极限定理来揭示。中心极限定理指出: 如果总体变量存在有限的平均数和方差, 那么不论这个总体变量的分布如何, 随着抽样单位数的增加, 抽样平均数便趋近于正态分布。这个结论对于抽样推断是十分重要的, 同时为抽样误差的概率估计提供了一个极为有效且方便的条件。

而参数估计和假设检验的区别在于: 参数估计是对总体数量特征数值的估算, 假设检验是对总体是否具有某种性质的判断。例如, 根据 200 件产品 (样本) 的合格率推算全部产品的合格率, 属于参数估计; 而根据技术改造后的 200 件产品 (样本) 的合格率判断技术改造是否有效, 则属于假设检验。

## 二、总体参数的估计

总体参数的估计方法有点估计和区间估计两种。点估计是根据样本统计量估计总体参数是什么点值, 区间估计是根据样本统计量估计总体参数所在的范围。

### 1. 点估计

根据样本统计量直接估计出总体参数  $\theta$  的值, 称为参数的点估计。

点估计常用的方法有两种: 矩估计法和极大似然估计法。

(1) 矩估计法由英国统计学家皮尔逊提出。其基本思想是: 由于样本来源于总体, 样本矩在一定程度上反映了总体矩, 而且由大数定律可知, 样本矩依概率收敛于总体矩。因此, 只要总体的  $k$  阶原点矩存在, 就可以用样本矩作为相应总体矩的估计量, 用样本矩的函数作为总体矩的函数的估计量。例如, 用样本均值来估计总体均值, 用样本方差来估计总体方差。

矩估计法简单、直观, 而且不必知道总体的分布类型, 因而得到了广泛应用。

按照矩估计法, 可以用样本均值直接作为总体平均数的估计值, 用样本成数直接作为总体成数的估计值, 用样本方差直接作为总体方差的估计值, 等等。例如, 如果要估计一批灯泡的平均寿命, 根据一个样本计算的平均寿命是 5 600 小时, 我们就用 5 600 小时作为整批灯泡的平均寿命的一个估计值, 这就是点



估计。又如，我们要估计一批产品的合格率，根据抽样结果合格率为 98%，将 98% 直接作为这批产品合格率的估计值，这也是点估计。

矩估计法也有局限性，它要求总体以  $k$  阶原点矩存在，否则无法估计；它不考虑总体分布类型，因此也就没有充分利用总体分布函数提供的信息。

(2) 极大似然估计法是由费希尔提出的一种参数估计方法。其基本思想是：已知某个参数能使这个样本出现的概率最大，我们当然不会再去选择其他小概率的样本，所以就在这个参数作为估计的真实值。当然，极大似然估计只是一种粗略的数学期望，要知道它的误差大小还要做区间估计。

抽样误差的存在是必然的，但点估计不考虑误差范围，因此也就说明不了估计的准确性和可靠性。所以，尽管点估计推断出了总体参数的一个点值，但这个点值的准确性和可靠性却是无法衡量的。因此，点估计更主要的作用是为区间估计提供推断基础。

## 2. 区间估计

我们在用点估计值代表总体参数值的同时，还必须给出一个用于衡量点估计值可靠性的度量。也就是说，必须能说出点估计值与总体参数的真实值接近的程度。但遗憾的是，点估计难以做到这一点，因此需要进行区间估计。区间估计是在点估计的基础上，给出总体参数估计的一个范围。

总体参数的估计区间通常由样本统计量加减抽样误差而得到。与点估计不同，进行区间估计时，根据样本统计量的抽样分布，我们能够对样本统计量与总体参数的接近程度给出一个概率度量。下面以总体均值的区间估计为例说明区间估计的基本原理。

由样本均值的抽样分布可知，在重复抽样和无限总体抽样的情况下，样本均值的数学期望等于总体均值，即  $E(\bar{X}) = \mu$ ，样本均值的标准差为  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ 。

由此可以知道：样本均值  $\bar{x}$  落在总体均值  $\mu$  的两侧各为一个抽样标准差范围内的概率为 0.682 7，落在两个抽样标准差范围内的概率为 0.954 5，落在三个抽样标准差范围内的概率为 0.997 3，等等。

实际上，我们可以求出样本均值落在总体均值  $\mu$  的两侧任何一个抽样标准差范围内的概率。但实际估计时，情况恰恰相反。 $\bar{x}$  是已知的，而  $\mu$  是未知的，也正是我们要估计的。由于  $\bar{x}$  与  $\mu$  的距离是对称的，如果某个样本的平均值落

在  $\mu$  的两个标准差之内，反过来， $\mu$  也被包括在以  $x$  为中心左右两个标准差范围之内。因此，约有 95% 的样本均值会落在  $\mu$  的两个标准误差的范围内。也就是说，约有 95% 的样本均值所构造的两个标准误差的区间会包括  $\mu$ 。通俗地说，如果抽取 100 个样本来估计总体均值，由 100 个样本所构造的 100 个区间中约有 95 个区间包括总体均值，而另外 5 个区间则不包括总体均值。换句话说，从 100 个样本中任意抽取一个样本所构造的区间，有 95% 的可能性包含了总体均值  $\mu$ 。我们把 95% 称为置信水平或置信度，相应的区间称为置信区间。图 5-2 给出了区间估计的示意图。

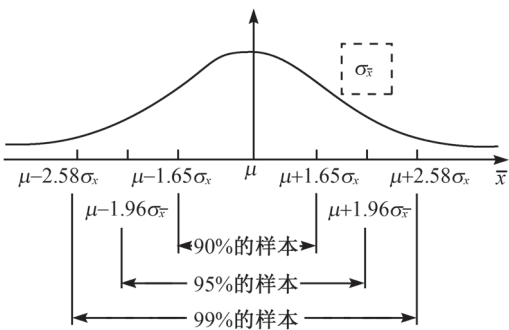


图 5-2 区间估计的示意图

具体来说，在区间估计中，由样本统计量所构造的总体参数的估计区间，称为置信区间，其中区间的最小值称为置信下限，最大值称为置信上限。

置信水平是指如果将构造置信区间的步骤重复多次，置信区间包含总体参数真值的次数所占的比例。

由此可见，总体均值  $\mu$  所在的区间取决于样本统计量的抽样分布，或者说，取决于置信水平。置信水平不同， $\mu$  所在的区间大小也不同。

在构造置信区间时，往往把所希望的某个值设为置信水平。常用置信水平的  $Z_{\alpha/2}$  值如表 5-3 所示。

表 5-3 常用置信水平的  $Z_{\alpha/2}$  值

| 置信水平 | $\alpha$ | $\alpha / 2$ | $Z_{\alpha/2}$ |
|------|----------|--------------|----------------|
| 90%  | 0.1      | 0.05         | 1.645          |
| 95%  | 0.05     | 0.025        | 1.96           |
| 99%  | 0.01     | 0.005        | 2.58           |

有关置信区间的概念可用图 5-3 表示。

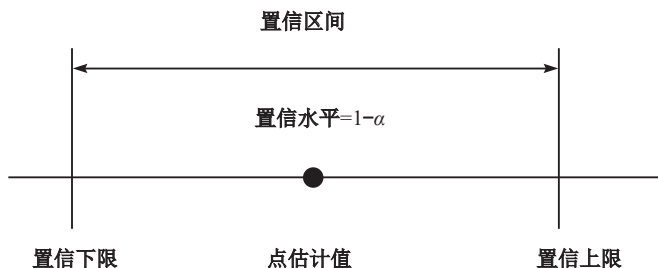


图 5-3 置信区间的示意图

从图 5-2 和图 5-3 可以看出, 当样本容量给定时, 置信区间的宽度随着置信水平的增大而增大。一般来说, 区间比较宽时, 才会使这一区间有更大的可能性包含参数的真值; 当置信水平固定时, 置信区间的宽度随样本容量的增大而减小, 换言之, 较大的样本所提供的有关总体的信息要比较小的样本多。

用一个具体的样本所构造的区间是一个特定的区间, 我们无法知道这个样本所产生的区间是否包含总体参数的真值, 只能希望这个区间是大量包含总体参数真值的区间中的一个, 但它也可能是少数几个不包含参数真值的区间中的一个。特别强调的是, 如果我们说总体参数以一定的概率 (如 95%) 落在这一区间, 这样的表述是错误的。例如, 从一个总体中抽取 20 个随机样本, 得到的总体均值  $\mu$  的 20 个估计区间, 每个区间中间的点表示  $\mu$  的点估计, 即样本均值, 可以看出 20 个区间中只有第 8 个区间没有包括总体均值  $\mu$ 。如果这是 95% 的置信区间, 最后只有 5% 的区间没有包含  $\mu$ , 如图 5-4 所示。

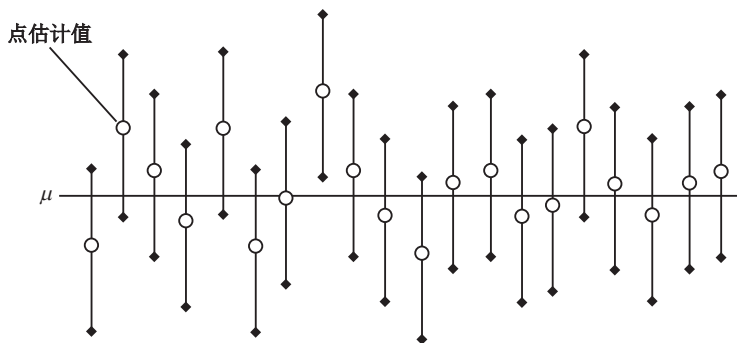


图 5-4 重复构造出的  $\mu$  的 20 个置信区间





### 任务卡 2

请写出你了解的确定样本容量的因素。

☐ 可接受的边际误差

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_

## 知识卡 2 样本容量的确定

### 一、估计总体均值时样本容量的确定

在实际抽样调查中,确定一个合适的样本容量是一个重要的问题。因为样本容量过多,必然会增加人力、财力、物力的支出,造成不必要的浪费;而样本容量过少,又会导致抽样误差增大,达不到抽样所要求的准确程度。

因此,必要的样本容量是在保证误差不超过规定范围的条件下尽可能节省人、财、物的支出。

前面已经讲过,总体均值的置信区间由样本均值和允许误差两部分组成。在重复抽样和无限总体抽样条件下,允许误差为  $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ 。临界值  $Z_{\alpha/2}$  和样本容量  $n$  共同确定了允许误差的大小。一旦确定了置信水平  $(1-\alpha)$ ,  $Z_{\alpha/2}$  的值就确定了。对于给定的  $Z_{\alpha/2}$  值和总体标准差  $\sigma$ ,就可以确定任意希望的允许误差所需要的样本容量。令  $E$  代表使用者在给定的置信水平下可以接受的边际误差,即

$$E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (5-5)$$

由此可以推导出确定样本容量的公式为

$$n = \frac{(Z_{\alpha/2})^2 \sigma^2}{E^2} \quad (5-6)$$

式中的  $Z_{\alpha/2}$  的值可直接由区间估计中所用到的置信水平确定。如果能求出



$\sigma$  的具体值, 就可以用式 (5-6) 计算所需的样本容量。在实际应用中, 如果  $\sigma$  的值未定, 可以用以前相同或类似的样本的标准差来代替, 也可以用实验调查的方法选择一个初始样本, 以该样本的样本标准差作为  $\sigma$  的估计值。

由式 (5-6) 可以看出, 样本容量与常用置信水平的  $Z_{\alpha/2}$  值的平方成正比, 在其他条件不变的情况下, 常用置信水平的  $Z_{\alpha/2}$  的值越大, 所需的样本容量也就越大; 样本容量与总体方差成正比, 总体的差异越大, 所要求的样本容量也越大; 样本容量与边际误差的平方成反比, 即可以接受的边际误差的平方越大, 所需的样本容量就越小。

需要说明的是, 根据式 (5-6) 计算出来的样本容量不一定是整数, 通常是将样本容量取成较大的整数, 也就是小数点后面的数值一律进位成整数, 如 36.78 取 37, 36.28 也取 37, 注意这里不能采用四舍五入法, 这就是样本容量的圆整法则。



想一想

## 二、估计总体比例时样本容量的确定

与估计总体均值时样本容量的确定方法类似, 在重复抽样和无限总体抽样条件下, 估计总体比例置信区间的允许误差为

$$Z_{\alpha/2} \sqrt{\frac{\pi(1-\pi)}{n}}$$

$Z_{\alpha/2}$  的值、总体比例  $\pi$  和样本容量  $n$  共同确定了边际误差的大小。一旦确定了置信水平  $(1-\alpha)$ ,  $Z_{\alpha/2}$  的值就确定了。由于总体比例的值是固定的, 允许误差由样本容量来确定, 样本容量越大, 允许误差就越小, 估计的精度就越好。因此, 对于给定的  $Z_{\alpha/2}$  值, 就可以确定任意希望的允许误差所需要的样本容量。令  $E$  代表所希望达到的允许误差, 即

$$E = Z_{\alpha/2} \sqrt{\frac{\pi(1-\pi)}{n}} \quad (5-7)$$

由此可以推导出重复抽样和无限总体抽样条件下确定样本容量的公式为

$$n = \frac{(Z_{\alpha/2})^2 \pi(1-\pi)}{E^2}$$

式中的允许误差  $E$  必须是使用者事先确定的。大多数情况下, 取  $E$  的值小



于 0.10。 $Z_{\alpha/2}$  的值可直接由区间估计中所用到的置信水平确定。如果能求出  $\pi$  的具体值, 就可以用上面的公式计算所需的样本容量。在实际应用中, 如果不知道  $\pi$  的值, 可以用类似的样本比例来代替; 也可以用试验调查的方法, 选择一个初始样本, 以该样本的比例作为  $\pi$  的估计值, 通常取使  $\pi(1-\pi)$  为最大值的 0.5。



想一想



素质园地



## 职业基础知识测试

### 一、填空题

1. 抽样调查是非全面调查, 它是按\_\_\_\_\_原则从调查对象(即总体)中抽取\_\_\_\_\_单位作为样本进行调查, 用调查所得指标数值对\_\_\_\_\_做出具有一定可靠性的估计和推断的一种统计调查方法。
2. 抽样技术具有\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_等优点。
3. 总体也称全及总体, 指所要认识的\_\_\_\_\_的全体。样本又称抽样总体或子样, 它是从总体中\_\_\_\_\_的, 作为代表总体\_\_\_\_\_的那部分单位组成的\_\_\_\_\_。
4. 抽样估计是建立在概率论的\_\_\_\_\_基础之上的。
5. 置信水平是指如果将\_\_\_\_\_的步骤重复多次, 置信区间包含的\_\_\_\_\_次数所占的比率。

### 二、简答题

1. 简述抽样调查的种类。
2. 抽样调查具有哪些作用?
3. 什么是抽样误差?
4. 影响抽样误差的因素有哪些?
5. 简述点估计。



## 职业核心能力训练

### 【实训目标】

了解抽样分布的概念、抽样分布的形成过程，力求寻找出样本均值的规律。

### 【具体要求】

某家庭有五口人，年龄分别为 60 岁、58 岁、32 岁、30 岁、5 岁。现从中随机抽取 2 人，观察其平均年龄的概率分布。

- (1) 列出所有可能样本的结果。
- (2) 计算出所有可能样本下的平均年龄及其概率。
- (3) 计算平均年龄概率分布的数学期望及方差。

### 【实训记录】

### 【交流与评价】

| 项 目  | 成 绩                     |                         |                      |                 |     |
|------|-------------------------|-------------------------|----------------------|-----------------|-----|
|      | 所有可能样本结果的全面性<br>( 20% ) | 平均年龄及其概率的正确性<br>( 20% ) | 期望与方差的正确性<br>( 25% ) | 最终结果<br>( 35% ) | 总 分 |
| 自我评价 |                         |                         |                      |                 |     |
| 同学评价 |                         |                         |                      |                 |     |
| 教师评价 |                         |                         |                      |                 |     |





## 项目六 熟悉时间序列分析

### 知识目标 →

- 了解时间序列的含义和作用，理解时间序列的分类，了解编制时间序列应注意的问题
- 理解发展水平与平均发展水平的概念
- 掌握增长量、平均增长量、发展速度、平均发展速度、增长速度、平均增长速度的概念

### 能力目标 →

- 能够编制时间序列
- 能够运用时间序列的水平分析服务于统计工作

### 思维导图 →

#### 熟悉时间序列分析

##### 认识时间序列

时间序列的含义、作用和种类

编制时间序列的原则

##### 掌握时间序列的水平分析

发展水平

增长量和平均增长量

##### 掌握时间序列的速度分析

发展速度和增长速度

平均发展速度和平均增长速度

##### 了解长期趋势分析和季节变动分析

长期趋势分析

季节变动分析



### 项目情境 →

小迪作为人力资源管理专业的应届毕业生入职某单位，人力总监每天会向她了解出勤情况，于是她单独列出某星期每天出勤的职工人数：300，320，340，330，320。在汇报过程中，小迪发现其实这些出勤人数可以形成时间序列，从而快速计算出平均每天出勤的职工人数。假如你是小迪，你如何把每天的出勤人数形成时间序列汇报给人力总监？

---

---

---

---





## 任务一 认识时间序列

小明：这学期的课程快完了，爸爸公司的新产品研发却在如火如荼地进行着。想想这些日子我跟着爸爸接触到这么多统计学方面的知识，虽然牺牲了不少课余时间，但还是感到由衷的高兴。现在很多同学有统计学方面的问题都问我，我也乐意帮他们解答。不过，接下来的这些课程也马虎不得。

### 任务卡 1

请你写出你所了解的时间序列的种类。

☐ 绝对数时间序列

☐ \_\_\_\_\_

☐ 相对数时间序列

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_

### 知识卡 1 时间序列的含义、作用 and 种类

#### 一、时间序列的含义和作用

时间序列又称时间数列或动态数列，是将某一指标在不同时间上的数值，按时间（如年、季、月等）先后顺序排列而成的统计数列。例如，将历年我国国民生产总值加以排列，或者将某企业某年每月的产品产量、职工人数、劳动生产率加以排列，都可以形成时间序列，如表 6-1 所示。

一般来说，时间序列是由两个基本数列构成的：一是现象所属时间的数列，如表 6-1 中的“2020”“2021”等；二是反映客观现象各个具体指标数值的数列，如表 6-1 中的每年的“生产总值”“年末人口数”等。

表 6-1 2017—2021 年某地区经济及人口统计

| 年 份  | 某地区生产总值<br>/ 万元 | 年末人口数 / 人 | 农村居民人均纯<br>收入 / 元 | 年末该地区就业<br>人数 / 人 |
|------|-----------------|-----------|-------------------|-------------------|
| 2017 | 112 398         | 108 453   | 2 376             | 70 740            |
| 2018 | 136 094         | 109 227   | 2 602             | 74 402            |
| 2019 | 137 620         | 129 908   | 2 906             | 75 400            |
| 2020 | 182 021         | 130 056   | 3 205             | 75 025            |
| 2021 | 209 007         | 131 048   | 3 507             | 76 000            |

时间序列在统计分析中，可以描述社会经济现象的发展情况、变化过程及内在规律，提供反映社会经济现象发展速度和变化规律的数据，计算相应指标体系，揭示现象发展变化的趋势，从而为统计预测提供依据。例如，在表 6-1 中生产总值等指标在 2017—2021 年期间表现为逐年增长的趋势。可见，时间序列是观察分析事物的一种重要方法，在社会经济现象动态分析中有着十分重要的作用。其主要作用如下：

（1）时间序列可以表明社会经济现象的发展变化趋势及规律性。例如，把相邻几年各季空调的销售量编制成时间序列，通过比较不仅会发现空调的销售量有不断增长的趋势，而且会发现销售量的季节变动规律。

（2）可以根据时间序列计算各种时间动态指标值，以便具体、深入地揭示现象发展变化的数量特征。

（3）运用时间序列可以预测现象的发展方向和发展速度，为经济决策或经营决策提供重要依据。

总之，通过时间序列的编制和分析，可以研究过去、指导现在和预测未来。

二、时间序列的种类

时间序列根据其指标表现形式可分为总量指标时间序列、相对指标时间序列和平均指标时间序列。在三种时间序列中，总量指标时间序列是最基本的原始序列，而相对指标时间序列和平均指标时间序列则是在其基础上计算产生的，称为派生数列。

1. 总量指标时间序列

总量指标时间序列，是由一系列同类总量指标的数值按时间的先后次序排



列而成的时间序列。它反映的是社会经济现象的总量在各个时期所达到的规模、水平及发展变化的情况。总量指标时间序列根据其总量指标所反映的现象总量性质、时间状况的不同,又可分为时期序列和时点序列。

时期序列是由时期总量指标编制的时间序列,它反映了某种现象总体在一段时间内发展过程的总量。时间序列中的指标都是时期指标。例如,表 6-1 中某地区的生产总值时间序列,由于生产总值是时期指标,因此该时间序列为时期序列。时期序列中的各项指标所涉及的时间长度称为“时期”,指标数值的大小与时期的长短有直接关系。时期越长,数值越大;时期越短,数值越小。因此,时期序列具有可加性,相加的结果反映现象在更长时间内的的发展总量。

时点序列是用时点总量指标编制的时间序列,序列中的每项指标数值都是反映现象总体在某一时刻或时点上所达到的总量水平。例如,表 6-1 中的年末人口数、年末该地区就业人数时间序列都是时点序列。在时点序列中,两个相邻时点之间的时间间隔称为“时点间隔”。它可以是相等的,也可以是不等的。表 6-1 中的时点间隔是相等的,均为一年。时点序列中的指标数值与时点间隔的长短无直接的联系,指标数值是现象在一段时间内增减抵消后的结果。因此,时点序列不具有可加性。

## 2. 相对指标时间序列

相对指标时间序列,是由不同时间上同类相对指标数值按时间先后顺序排列而成的序列。它反映社会经济现象之间的数量对比关系或说明现象的结构、速度等的发展变化过程。表 6-2 所示的某地区第一产业比重资料就是相对指标时间序列。

表 6-2 某地区第一产业比重资料

| 年 份              | 2013  | 2014  | 2015  | 2016  | 2017  | 2018  | 2019  | 2020  | 2021  |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 第一产业占某地区的生产总值的比重 | 26.1% | 20.5% | 22.8% | 19.8% | 20.3% | 20.5% | 20.4% | 19.2% | 18.4% |

由于相对指标一般表现为两个绝对指标之比,因此相对指标时间序列可以看作是由总量指标时间序列派生出来的。



### 3. 平均指标时间序列

平均指标时间序列是由不同时间上同类平均指标按时间先后顺序排列而成的统计数列。它可以用来反映各个时期社会经济现象一般水平的发展过程和变化趋势。除平均工资以外,平均亩产量、工人劳动生产率等平均指标所构成的时间序列,也都是平均指标时间序列。

平均指标一般也表现为两个绝对指标之比,因此,平均指标时间序列也可以看作是由总量指标时间序列派生出来的。

#### 任务卡 2

你了解编制时间序列的原则吗? 查找资料,写下你的收获。

### 知识卡 2 编制时间序列的原则

时间序列的动态分析是通过同一指标不同时间的数值对比,来反映社会经济现象的发展变化过程及规律性。因此,保证时间序列中各时间指标数值的可比性,是正确编制时间序列应遵守的基本原则。编制时间序列应遵守以下具体原则。

#### 1. 时间长短的可比性

如果时间序列中的各项指标都是时期指标,其数值所属的时期长短应该相等。因为时期指标的数值大小与时间长短有直接关系,只有时期长短保持一致,才能保证各项指标值之间具有可比性。如果时间序列中的各项指标都是时点序列,其数值之间的时间间隔也尽量相等,以便对比分析。在特殊情况下,也可以将不同时间长度的同类指标组成的时间序列进行比较分析,以强调说明特殊问题,如表 6-3 所示。



表 6-3 某地区几个重要时期的小麦产量统计

单位: 万吨

| 年份 | 1988—1997 | 1998—2003 | 2004—2009 | 2010—2015 | 2016—2021 |
|----|-----------|-----------|-----------|-----------|-----------|
| 产量 | 876       | 10 230    | 17 560    | 35 640    | 67 800    |

表 6-3 中的第一项指标是某地区以前 10 年小麦产量的总和, 后四项指标为此地区以后各六年计划的实际小麦产量, 虽然两者的时间长度不相等, 但其目的在于说明此地区经济的迅速发展情况。此例旨在说明时间序列中时间要求的灵活性。但就一般情况而言, 仍应该使序列中各项指标的时间长度相等。

## 2. 总体范围的可比性

时间序列所属的总体范围应相同, 若被研究现象所属空间范围发生变动, 引起有关时间序列指标的变化, 使其总体范围前后不一致, 则所形成的时间指标不可比。所以, 在同一时间序列中, 各个指标所包括的总体范围前后应一致。例如, 如果研究某个地区总人口数的变化情况, 就必须保证该地区前后有相同的管辖范围。如果管辖范围发生了变化, 那么总人口数资料也应该做相应的调整。

## 3. 经济内容的可比性

时间序列中的指标, 有时会出现名称相同、但其经济内容不完全相同的情况, 如果不注意, 就会影响我们对问题的分析。所以, 编制时间序列时, 还应注意历史条件的变化所引起的指标经济内容和性质的变化。不同性质的指标不能编制同一时间序列。特别是研究不同的社会制度, 或者研究经济大变革时期的经济发展变化, 更要注意这个问题。例如, 商品价格有购进价格和销售价格之分, 如果把这两种价格混在一起构成时间序列, 就会导致得出错误的分析结论。

## 4. 计算方法的一致性

时间序列中各项指标的计算方法、计量单位和计价标准前后要一致, 并在一定时期内保持不变。有时因计算方法不一致, 在价值指标中计算价格不统一, 在实物指标中计量单位不统一等, 在指标数值上也就不具有可比性。例如, 要研究企业人均销售量的变动, 产量用实物量还是用价值量, 人数用全部职工人数还是销售人员数, 前后都要统一起来。这样对比的结果才有意义, 才有使用价值。



## 任务二 掌握时间序列的水平分析

### 任务卡 1

统计学的“发展水平”和经济上的“发展水平”的含义一样吗？在课堂进行讨论，说说你的看法。

### 知识卡 1 发展水平

发展水平又称发展量，是指时间序列中每一项具体的统计指标数值。它具体反映社会经济现象在各个不同时期或时点上所达到的规模 and 水平，通过不同时期发展水平的比较，可以给人具体的、深刻的印象。

作为发展水平，它既可以是总量指标，也可以是相对指标或平均指标。由总量指标组成的时间序列，其指标数值即总量指标发展水平；由相对指标组成的时间序列，其指标数值即相对指标发展水平；由平均指标组成的时间数列，其指标数值即平均指标发展水平。

由于发展水平指标在时间序列中所处的位置不同，发展水平可以分为最初水平、最末水平和中间水平。处于时间序列首项的指标数值称为最初水平，处于末项的指标数值称为最末水平，中间各项的指标数值称为中间水平。如果用符号  $a_0, a_1, a_2, \dots, a_{n-1}, a_n$  代表时间序列中的各个发展水平，则  $a_0$  是最初水平， $a_n$  是最末水平， $a_1, a_2, \dots, a_{n-1}$  是中间水平。

在利用时间序列指标进行对比分析时，常把所研究那个时间的发展水平称为报告期水平或计算期水平，而把选做对比基础的发展水平称为基期水平。无





论哪一种发展水平，都不是固定不变的，而是随着研究目的的变化而变化。也就是说，现在的报告期水平，可能是将来的基期水平；这个序列的最末水平，也可能为另一个序列的最初水平。

应当注意的是，发展水平指标在文字叙述上习惯用“增加到”“增加为”或“降低到”“降低为”表示。例如，2020年某地区生产总值为300 670万元，2021年增加到335 353万元。又如，2020年某工厂某种产品的单位成本为32元，2021年降低到25元，等等。运用时，一定不要把“到”和“为”字漏掉，否则要说明的社会经济现象指标的意义就要发生变化。

平均发展水平是指时间序列中不同时期的发展水平采用一定的方法加以加权平均求得的平均数。它表明现象在一段时间内发展水平达到的一般水平，是根据序列中不同时期（或时点）上的发展水平计算的平均数。由于它是将社会经济现象在不同时期上的数量差异平均化而求得的，为了与前面学过的平均数有所区别，通常又把它称为序时平均数或动态平均数。

平均发展水平和静态平均数有相同之处，两者都是将现象的个别数量差异抽象化，所得到的现象的一般水平；但也存在区别，主要表现为静态平均数是根据变量数列计算的，而动态平均数是根据动态数列计算的。静态平均数是将总体各单位在同一时间上的数量差异抽象化、从时间截面上反映总体的一般水平，而动态平均数是将总体在不同时间上的数量差异抽象化、从时间过程上反映总体的一般水平。由于两类平均数的性质不同，因此其计算方法自然也不相同。

平均发展水平在动态分析中具有重要的意义，它可以把时间长短不等的问题指标由不可比变为可比，并消除现象在短期内波动的影响，便于观察现象的发展变化趋势和规律性。

按照时间序列的类型不同，可把平均发展水平的计算分为三种情况来计算。

### 一、总量指标时间序列序时平均数的计算

总量指标时间序列分时期序列和时点序列，由于两种序列所具有的性质不同，因而在计算序时平均数时，方法上也不一样。

#### 1. 依据时期序列计算序时平均数

由于时期序列中各项指标数值可以加总，加总的结果反映现象在较长时间





内发展变化的总量，因此它的序时平均数可以用简单算术平均方法计算，即将序列中各项指标的数值直接加总除以发展水平的项数。用公式表示为

$$\bar{a} = \frac{a_1 + a_2 + a_3 + \cdots + a_n}{n} = \frac{\sum a_i}{n}$$

式中， $\bar{a}$  为平均发展水平； $a_i$  为各期发展水平； $n$  为时期指标项数。

【例 6-1】某商场 2021 年各月商品销售额动态资料如表 6-4 所示，试计算月平均销售额。

表 6-4 某商场 2021 年各月商品销售额动态资料

| 月 份 | 销售额 / 万元 | 月 份  | 销售额 / 万元 |
|-----|----------|------|----------|
| 1 月 | 100      | 7 月  | 140      |
| 2 月 | 110      | 8 月  | 130      |
| 3 月 | 120      | 9 月  | 150      |
| 4 月 | 120      | 10 月 | 160      |
| 5 月 | 110      | 11 月 | 150      |
| 6 月 | 130      | 12 月 | 170      |

解：商品销售额是时期指标，由于各月商品销售额高低不等，因而发展变化趋势不够明显。如果计算出各季的月平均销售额，就会明显地反映出销售趋势。例如：

$$\text{第一季度月平均销售额：}\bar{a} = \frac{\sum a_i}{n} = \frac{100+110+120}{3} = 110 \text{（万元）}$$

$$\text{第二季度月平均销售额：}\bar{a} = \frac{\sum a_i}{n} = \frac{120+110+130}{3} = 120 \text{（万元）}$$

$$\text{第三季度月平均销售额：}\bar{a} = \frac{\sum a_i}{n} = \frac{140+130+150}{3} = 140 \text{（万元）}$$

$$\text{第四季度月平均销售额：}\bar{a} = \frac{\sum a_i}{n} = \frac{160+150+170}{3} = 160 \text{（万元）}$$

$$\text{全年月平均销售额：}\bar{a} = \frac{\sum a_i}{n} = \frac{100+110+\cdots+170}{12} = 132.5 \text{（万元）}$$

可见，该商场 2021 年的第三、四季度的月平均销售额大于第一、第二季度的月平均销售额。



## 2. 依据时点序列计算序时平均数

时点序列有连续时点序列和间断时点序列两种,这两种时点序列又有间隔相等和间隔不等两种表现形式。

如果时点序列中的数值是逐日记录、逐日排列的,就称为连续时点序列,否则就称为间断时点序列。

(1) 连续时点序列的序时平均数。连续时点序列是按日登记取得资料的时点数列,其序时平均数采取如下公式来计算:

$$\bar{a} = \frac{\sum a}{n}$$

式中,  $\bar{a}$  为每天的时点水平;  $n$  为天数。

用这种方法计算的序时平均数,其结果是非常精确的,但在实际工作中往往不可能或不必要取得每日的时点资料,因此,这种方法在实际运用上有很大的局限性。

如果连续时点序列的数据整理成了分组资料,则也可借助加权算术平均的形式来计算序时平均数。

【例 6-2】某企业第一季度出勤的职工人数变动情况如表 6-5 所示。试计算该企业第一季度平均职工人数。

表 6-5 某企业第一季度出勤的职工人数变动情况

|          |       |       |       |       |       |
|----------|-------|-------|-------|-------|-------|
| 职工人数 / 人 | 1 250 | 1 264 | 1 270 | 1 272 | 1 280 |
| 出勤天数 / 天 | 20    | 19    | 20    | 15    | 16    |

解: 由于第一季度全部 90 天的职工人数资料都齐全,因此本例属于连续时点序列。则该企业第一季度平均每日的职工人数为

$$\begin{aligned}
 \bar{a} &= \frac{\sum af}{f} \\
 &= \frac{1\,250 \times 20 + 1\,264 \times 19 + 1\,270 \times 20 + 1\,272 \times 15 + 1\,280 \times 16}{20 + 19 + 20 + 15 + 16} \\
 &= \frac{113\,976}{90} \approx 1\,266 \text{ (人)}
 \end{aligned}$$

(2) 间断时点序列的平均发展水平。间断时点序列是指按月末、季末或年末登记取得资料的时点序列。它有两种情况:一是序列中的各项指标表现为逐



想一想



期期末登记排列，二是序列中的各项指标表现为非均衡的期末登记排列。通常将前者称为间隔相等的间断时点序列，后者称为间隔不等的间断时点序列。

如果掌握的时点序列是时间间隔的期初或期末资料，则要以简单算术平均法分层计算其序时平均数，则间隔相等的间断时点序列的平均发展水平的计算公式为

$$\bar{a} = \frac{\frac{a_1 + a_2}{2} + \frac{a_2 + a_3}{2} + \cdots + \frac{a_{n-1} + a_n}{2}}{n-1} = \frac{\frac{a_1}{2} + a_2 + \cdots + a_{n-1} + \frac{a_n}{2}}{n-1}$$

【例 6-3】某商业银行 2021 年 6 月至 9 月末现金库存额如表 6-6 所示。根据表中资料，计算该银行第三季度月平均现金库存额。

表 6-6 某商业银行 2021 年 6 月至 9 月末现金库存额 单位：万元

| 时间    | 6 月末 | 7 月末 | 8 月末  | 9 月末 |
|-------|------|------|-------|------|
| 现金库存额 | 700  | 900  | 1 000 | 900  |

解：第三季度的月平均现金库存额为

$$\begin{aligned}\bar{a} &= \frac{\frac{a_1 + a_2}{2} + \frac{a_2 + a_3}{2} + \cdots + \frac{a_{n-1} + a_n}{2}}{n-1} = \frac{\frac{a_1}{2} + a_2 + \cdots + a_{n-1} + \frac{a_n}{2}}{n-1} \\ &= \frac{\frac{700}{2} + 900 + 1\,000 + \frac{900}{2}}{4-1} = 900 \text{ (万元)}\end{aligned}$$

如果掌握的时点序列是时间间隔不相等的期初或期末资料，则要以时间间隔长度为权数对时点资料用加权算术平均法计算平均发展水平。利用这种方法计算平均发展水平，假设资料欠缺的时间里，现象的发展变化是均匀的，但实际情况并非如此，因此得出的结果是近似结果。

间隔不相等的间断时点序列的平均发展水平的计算公式为

$$\bar{a} = \frac{\frac{a_1 + a_2}{2} f_1 + \frac{a_2 + a_3}{2} f_2 + \cdots + \frac{a_{n-1} + a_n}{2} f_{n-1}}{f_1 + f_2 + \cdots + f_{n-1}}$$

式中， $a_i$  为时点水平； $f_i$  为两个相邻的时点之间的时间间隔长度（ $i=1, 2, \cdots, n-1$ ）。



【例 6-4】某城市 2021 年的外来人口资料如表 6-7 所示。计算该市 2021 年平均外来人口数。

表 6-7 某城市 2021 年的外来人口资料

单位：万人

| 时间    | 1 月 1 日 | 5 月 1 日 | 8 月 1 日 | 12 月 31 日 |
|-------|---------|---------|---------|-----------|
| 外来人口数 | 21.30   | 21.38   | 21.40   | 21.51     |

解：该市 2021 年平均外来人口数为

$$\begin{aligned}\bar{a} &= \frac{\frac{a_1 + a_2}{2} f_1 + \frac{a_2 + a_3}{2} f_2 + \cdots + \frac{a_{n-1} + a_n}{2} f_{n-1}}{f_1 + f_2 + \cdots + f_{n-1}} \\ &= \frac{\frac{21.3 + 21.38}{2} \times 4 + \frac{21.38 + 21.4}{2} \times 3 + \frac{21.4 + 21.51}{2} \times 5}{4 + 3 + 5} \\ &= \frac{256.805}{12} \approx 21.4 \text{ (万人)}\end{aligned}$$

在对间断时点序列计算序时平均数时要注意， $a_1$  必须是所要平均的时间段的期初资料，而  $a_n$  必须是该时间段的期末资料。例如，在例 6-3 中，要计算第三季度的平均数，则必须把 6 月末的资料 700 万元作为  $a_1$ ，而不能把 7 月末的资料作为  $a_1$ ，因为 6 月末的资料恰恰是第三季度的期初资料。同理，应该把 9 月末（第三季度期末）的资料作为  $a_n$ ，只有这样才能计算出第三季度全期的平均数。

## 二、相对指标时间序列序时平均数的计算

相对指标时间序列是由两个相互联系的时间序列对比而求得的，而且分子、分母两个指标的时间状况一般不相同，因此要分别计算出分子、分母两个总量指标时间序列的序时平均数，而后加以对比来求得相对指标时间序列的序时平均数。

相对指标时间序列的序时平均数的计算公式为

$$\bar{c} = \frac{\bar{a}}{\bar{b}}$$



式中,  $\bar{c}$  为相对指标时间序列的序时平均数;  $\bar{a}$  为分子序列的序时平均数;  $\bar{b}$  为分母序列的序时平均数。

由于相对指标时间序列可由两个时期序列、两个时点序列或由一个时期序列和一个时点序列对比形成, 但需要注意时期序列与时点序列的平均发展水平的计算方法是不同的。



想一想

### 三、根据平均指标时间序列计算序时平均数

平均指标时间序列序时平均数的计算, 可分为如下两种情况:

(1) 由一般平均数组成的平均指标时间序列计算序时平均数。此时, 平均指标时间序列类似于上面所介绍的相对指标时间序列, 故可采用如下公式计算:

$$\bar{c} = \frac{\bar{a}}{\bar{b}}$$

【例 6-5】某企业 2021 年第三季度工人人数和总产值资料如表 6-8 所示。

表 6-8 某企业 2021 年第三季度工人人数和总产值资料

| 时 间      | 7 月   | 8 月   | 9 月   | 10 月  |
|----------|-------|-------|-------|-------|
| 工人人数 / 人 | 1 850 | 2 000 | 2 050 | 2 100 |
| 总产值 / 万元 | 260   | 290   | 310   | 300   |

计算该企业第三季度月平均劳动生产率。

解: 第三季度月平均总产值为

$$\bar{a} = \frac{260 + 290 + 310}{3} \approx 286.67 \text{ (万元)}$$

第三季度平均工人人数为

$$\bar{b} = \frac{\frac{1850}{2} + 2000 + 2050 + \frac{2100}{2}}{4-1} \approx 2008 \text{ (人)}$$

所以, 第三季度月平均劳动生产率为

$$\bar{c} = \frac{\bar{a}}{\bar{b}} = \frac{286.67}{2008} \approx 0.1428 \text{ (万元)}$$

(2) 由序时平均数组成的平均指标时间序列计算平均发展水平。如果时间序列中各个时期的间隔相等, 可用简单算术平均法计算序时平均数; 如果数列中



各个时期的间隔不等,则以间隔长度为权数,用加权算术平均法计算平均发展水平。

$$\bar{c} = \frac{\sum \bar{a}}{n}$$

或

$$\bar{c} = \frac{\sum \bar{a}f}{\sum f}$$

【例 6-6】某百货大楼 2021 年某商品各季度月均销售额如表 6-9 所示。

表 6-9 某百货大楼 2021 年某商品各季度月均销售额

| 时 间         | 一季度 | 二季度 | 三季度 | 四季度 |
|-------------|-----|-----|-----|-----|
| 月平均销售额 / 万元 | 150 | 200 | 180 | 210 |

试计算全年的平均月销售额。

$$\text{解: } \bar{c} = \frac{\sum \bar{a}}{n} = \frac{150+200+180+210}{4} = 185 \text{ (万元)}$$

### 任务卡 2

小明家里种了 20 亩 (1 亩  $\approx$  666.67 平方米) 大豆,去年的总产量为 5 000 千克,但是今年由于发生洪水,相较于去年增长量为 -300 千克。对此,你认为他描述的增长量为负数,这种说法是正确的吗?说说你的看法。

---



---

## 知识卡 2 增长量和平均增长量

### 一、增长量

增长量又称增减量,是指在一定时期内发展水平增减的绝对量,即时间序列中报告期水平与基期水平之差,说明社会经济现象在一定时期内增减变化的



绝对量。其计算公式为

$$\text{增长量} = \text{报告期水平} - \text{基期水平}$$

增长量指标可正可负。如果计算结果为正值，则为增长的绝对量；如果计算结果为负值，则表示减少或降低的绝对量。

在计算增长量时，由于研究的目的不同，选择的基期也不同。通常增长量指标可分为逐期增长量和累计增长量。

逐期增长量又称环比增长量，是报告期水平与前期水平之差，表明报告期较前期增减变化的绝对量，用符号表示则为

$$a_1 - a_0, a_2 - a_1, a_3 - a_2, \cdots, a_n - a_{n-1}$$

累计增长量又称定基增长量，是报告期水平与某一固定基期水平（通常为最初水平）之差，表明报告期较某一固定基期增减变化的绝对量，用符号表示则为

$$a_1 - a_0, a_2 - a_0, a_3 - a_0, \cdots, a_n - a_0$$

这两种增长量虽然是分别根据不同的基期计算的，但它们之间存在一定的联系，这种联系具体表现为累计增长量等于相应的各个逐期增长量之和，逐期增长量等于相邻的两个累计增长量之差，用符号表示则为

$$a_n - a_0 = (a_1 - a_0) + (a_2 - a_1) + (a_3 - a_2) + \cdots + (a_n - a_{n-1})$$

$$a_n - a_{n-1} = (a_n - a_0) - (a_{n-1} - a_0)$$

表 6-10 为某企业 2017—2021 年商品销售总额的逐期增长量与累计增长量。

表 6-10 某企业 2017—2021 年商品销售总额的逐期增长量与累计增长量

| 年 份      |    | 2017    | 2018    | 2019    | 2020     | 2021     |
|----------|----|---------|---------|---------|----------|----------|
| 销售额 / 万元 |    | 7 250.3 | 8 245.7 | 9 704.8 | 12 462.1 | 16 164.7 |
| 增长量 / 万元 | 逐期 |         | 995.4   | 1 459.1 | 2 757.3  | 3 802.6  |
|          | 累计 |         | 995.4   | 2 454.5 | 5 211.8  | 9 014.4  |

此外，在实际工作中，为了消除季节变动的影响，常计算年距增长量指标。年距增长量的计算公式为

$$\text{年距增长量} = \text{本期发展水平} - \text{去年同期发展水平}$$





## 二、平均增长量

平均增长量又称平均增减量,是指某一现象在一定时期内平均每期增减变化的数量,即逐期增长量的序时平均数,表明社会经济现象在一定时期内平均每期增长的数量。其计算方法是逐期增长量之和除以逐期增长量的个数,用公式表示为

$$\begin{aligned}\text{平均增长量} &= \text{逐期增长量之和} \div \text{逐期增长量的个数} \\ &= \frac{\text{累计增长量}}{\text{时间序列项数}} - 1\end{aligned}$$

仍以表 6-10 的资料来说明。根据表 6-10 提供的资料,该企业 2017—2021 年平均每年的增长量为

$$\begin{aligned}\bar{a} &= \frac{995.4 + 1\,459.1 + 2\,757.3 + 3\,802.6}{4} = \frac{9\,014.4}{4} \\ &= 2\,253.6 \text{ (万元)}\end{aligned}$$

或

$$\bar{a} = \frac{9\,014.4}{5-1} = \frac{9\,014.4}{4} = 2\,253.6 \text{ (万元)}$$

这表明,该企业这五年的销售额平均每年增长 2 253.6 万元。

## 任务三 掌握时间序列的速度分析

### 任务卡 1

请你写出你所了解的时间序列的内容。

- |                                 |                                |
|---------------------------------|--------------------------------|
| <input type="checkbox"/> 平均增长速度 | <input type="checkbox"/> _____ |
| <input type="checkbox"/> 发展速度   | <input type="checkbox"/> _____ |
| <input type="checkbox"/> _____  | <input type="checkbox"/> _____ |



## 知识卡 1 发展速度和增长速度

### 一、发展速度

发展速度是研究某种社会经济现象发展程度的动态分析指标。它是用时间序列中的报告期水平与基期水平之比来求得的，反映某种现象的发展方向和程度，一般用百分数表示；当发展速度较大时，也可以用倍数表示。

其一般计算公式为

$$\text{发展速度} = \frac{\text{报告期水平}}{\text{基期水平}} \times 100\%$$

当发展速度大于 100% 时，表示上升，小于 100% 时，表示下降。

由于采用的基期不同，发展速度指标可分为环比发展速度和定基发展速度。

环比发展速度是报告期水平与上一期水平之比，反映社会经济现象逐期发展变化的相对程度。其计算公式为

$$\text{环比发展速度} = \frac{\text{报告期水平}}{\text{报告期上一期水平}} \times 100\%$$

用符号表示为

$$\frac{a_1}{a_0}, \frac{a_2}{a_1}, \frac{a_3}{a_2}, \dots, \frac{a_n}{a_{n-1}}$$

定基发展速度是报告期水平与某一固定基期水平之比，反映社会经济现象在较长一段时间内总的发展变化程度，故又称总发展速度。其计算公式为

$$\text{定基发展速度} = \frac{\text{报告期水平}}{\text{某一固定基期水平}} \times 100\%$$

用符号表示为

$$\frac{a_1}{a_0}, \frac{a_2}{a_0}, \frac{a_3}{a_0}, \dots, \frac{a_n}{a_0}$$

例如，表 6-11 为某地区 2017—2021 年生产总值及其发展速度，根据此表可计算出各年的发展速度指标。



表 6-11 某地区 2017—2021 年生产总值及其发展速度

| 年 份       | 2017    | 2018    | 2019    | 2020    | 2021    |
|-----------|---------|---------|---------|---------|---------|
| 生产总值 / 万元 | 102 398 | 116 694 | 136 515 | 182 321 | 209 407 |
| 环比发展速度    |         | 113.96% | 116.99% | 133.55% | 114.86% |
| 定基发展速度    |         | 113.96% | 133.32% | 178.05% | 204.50% |

由于选择的基期不同, 定基发展速度与环比发展速度所反映现象发展变化的经济含义也不相同。然而, 它们之间存在一定的数量关系, 具体表现为:

(1) 定基发展速度等于相应的各个环比发展速度的连乘积, 即

$$\frac{a_n}{a_0} = \frac{a_1}{a_0} \cdot \frac{a_2}{a_1} \cdot \frac{a_3}{a_2} \cdot \dots \cdot \frac{a_n}{a_{n-1}}$$

(2) 两个相邻时期的定基发展速度之比, 等于相应的环比发展速度, 即

$$\frac{\frac{a_n}{a_0}}{\frac{a_{n-1}}{a_0}} = \frac{a_n}{a_0} \div \frac{a_{n-1}}{a_0}$$

为了消除季节因素变动对社会经济现象发展变化的影响, 在计算月份或季度发展速度时, 可选用上年同期作为对比的基期, 计算年距发展速度, 用以说明本期发展水平与去年同期发展水平对比而达到的相对发展程度。年距发展速度的计算公式为

$$\text{年距发展速度} = \frac{\text{本期发展水平}}{\text{去年同期发展水平}}$$

## 二、增长速度

增长速度又称增减速度, 是报告期增长量与基期发展水平之比。它反映社会经济现象在一定时期内增减程度的动态分析指标, 一般用百分数或倍数表示。其计算公式为

$$\begin{aligned} \text{增长速度} &= \frac{\text{增长量}}{\text{基期水平}} \times 100\% \\ &= \frac{\text{报告期水平} - \text{基期水平}}{\text{基期水平}} \times 100\% \\ &= \text{发展速度} - 1 \end{aligned}$$

增长速度指标可正可负。当发展速度大于 100%, 增长量为正值时, 则增长



速度为正数，表明为递增速度；当发展速度小于 100%，增长量为负值时，则增长速度为负数，表明为递减速度。

在计算增长速度时，由于采用的增长量和对比的基期水平不同，增长速度可分为定基增长速度和环比增长速度两种。

定基增长速度是报告期的累计增长量与某一固定基期的水平（通常为最初水平）之比，表明某种社会经济现象在较长一段时间内总的增长速度。其计算公式为

$$\begin{aligned}\text{定基增长速度} &= \frac{\text{累计增长量}}{\text{某一固定基期水平}} \\ &= \frac{\text{报告期水平} - \text{固定基期水平}}{\text{固定基期水平}} \\ &= \text{定基发展速度} - 1\end{aligned}$$

环比增长速度是报告期的逐期增长量与前一期发展水平之比，表明社会经济现象逐期的增长速度。其计算公式为

$$\begin{aligned}\text{环比增长速度} &= \frac{\text{逐期增长量}}{\text{前一期发展水平}} \\ &= \frac{\text{报告期水平} - \text{前一期发展水平}}{\text{前一期发展水平}} \\ &= \text{环比发展速度} - 1\end{aligned}$$

现根据表 6-11 提供的资料，计算 2017—2021 年某地区生产总值资料的增长速度指标，如表 6-12 所示。

表 6-12 2017—2021 年某地区生产总值及其增长速度

| 年 份       | 2017    | 2018    | 2019    | 2020    | 2021    |
|-----------|---------|---------|---------|---------|---------|
| 生产总值 / 万元 | 102 398 | 116 694 | 136 515 | 182 321 | 209 407 |
| 环比增长速度    |         | 13.96%  | 16.99%  | 33.55%  | 14.86%  |
| 定基增长速度    |         | 13.96%  | 33.32%  | 78.05%  | 104.50% |

从表 6-12 中计算的指标数值可知，2017—2021 年某地区生产总值是逐年增长的，但各年增长的速度大小不一样，有的年份相对高一些，有的年份相对低一些。



从上面的计算公式和所举例子可以看出,定基增长速度与环比增长速度之间不存在直接的换算关系,也就是环比增长速度的连乘积不等于定基增长速度。如需推算,必须将增长速度转化为发展速度,利用发展速度的关系互相推算,再转化为增长速度。

此外,在实际统计工作中,为了消除季节变动的影响,也常计算年距增长速度,用以说明年距增长量与去年同期发展水平对比达到的相对增长程度。年距增长速度的计算公式为

$$\text{年距增长速度} = \frac{\text{年距增长量}}{\text{去年同期发展水平}}$$

为了把速度指标、水平指标结合起来,深入分析环比增长速度与逐期增长量之间的关系,进一步反映增长速度的实际效果,有必要计算环比增长速度每增加一个百分点所代表的绝对量,通常称为增长1%的绝对量。其计算公式为

$$\text{增长1\%的绝对量} = \frac{\text{逐期增长量}}{\text{环比增长速度} \times 100} = \frac{\text{前一期水平}}{100}$$

社会经济现象在不同时期的发展速度是不同的。为了说明社会经济现象在一段较长时期内发展变化的一般程度,必须将社会经济现象在这个时期内的发展速度差异加以抽象,计算平均速度指标。平均速度指标有平均发展速度和平均增长速度两种。

### 任务卡 2

你了解平均发展速度和平均增长速度吗?查找资料,说说两者的区别。

## 知识卡 2 平均发展速度和平均增长速度

### 一、平均发展速度

平均发展速度和平均增长速度是动态分析的重要指标,在国民经济建设和



企业的经营管理中都有着十分重要的作用。例如,在编制五年计划或十年规划时,除规定发展水平外,还要计算每年的平均发展速度。在进行不同地区、部门及国家间经济发展情况的对比时,通过平均速度的比较,可以找出先进与落后间的差距,规划出今后的奋斗目标,加速经济发展。

由于社会经济现象发展的总速度不等于各年发展速度之和,而等于各年环比发展速度的连乘积,因此平均发展速度不能用算术平均法计算,而要用几何平均法计算,这种方法称为几何平均法或水平法。几何平均法计算平均发展速度时,是将各个环比发展速度视作变量( $x$ ),将环比发展速度的个数视作变量值的个数( $n$ )。计算公式为

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n} = \sqrt[n]{\prod x} \quad (6-1)$$

式中, $\bar{x}$ 为平均发展速度; $x$ 为各期环比发展速度; $\prod$ 为连乘符号; $n$ 为环比发展速度的项数。

由于时间序列中的定基发展速度等于各环比发展速度的连乘积,因此,式(6-1)可变化为

$$\bar{x} = \sqrt[n]{\frac{a_1}{a_0} \cdot \frac{a_2}{a_1} \cdot \frac{a_3}{a_2} \cdots \frac{a_n}{a_{n-1}}} = \sqrt[n]{\frac{a_n}{a_0}} \quad (6-2)$$

一段时期的定基发展速度即为现象的总速度。如果用 $R$ 表示总速度,则平均发展速度的计算公式还可以表示为

$$\bar{x} = \sqrt[n]{R} \quad (6-3)$$

上述三个计算平均发展速度的公式虽然形式不同,但其实质内容与结果是完全一样的,只是所运用的资料不同而已。究竟采用哪个公式,主要取决于所掌握的资料。利用几何平均法求现象的平均发展速度,可以借助相对数计算,也可以直接用多功能电子计算器计算。若已知各期的环比发展速度资料,可采用式(6-1)计算;若已知最初水平和最末水平,可采用式(6-2)计算;若给出了一个较长时期的总发展速度指标,则利用式(6-3)计算。

现根据表 6-11 提供的生产总值及发展速度资料,计算 2017—2021 年的国内生产总值的平均发展速度。

$$x = \sqrt[4]{113.96\% \times 116.99\% \times 133.55\% \times 114.86\%} \approx 119.58\%$$



或

$$x = \sqrt[4]{\frac{209\,407}{102\,398}} = \sqrt[4]{2.0450} \approx 119.58\%$$

平均增长速度 = 119.58% - 100% = 19.58%。

计算结果表明,用以上公式对同一现象计算平均发展速度,其计算结果相同(有时会出现小数不一致的情况,属于计算过程中四舍五入造成的误差)。但是这种方法不能准确反映中间水平的起伏状况。用几何平均法计算的这一指标,实质上只考虑数列的最初水平和最末水平,而没有考虑中间各项指标数值的影响,因此,运用几何平均法计算平均发展速度时,应特别注意以下问题:

(1) 社会经济现象发展变化的方向应基本保持一致。

(2) 要反映较长时期内社会经济现象总的平均发展速度时,还应与分段计算的平均发展速度及各期的实际发展水平、环比发展速度结合运用。

## 二、平均增长速度

平均发展速度是各时间环比发展速度的序时平均数,它说明社会经济现象在较长一段时间中各期平均发展变化的程度。平均增长速度则说明现象在较长一段时期中逐期平均增减变化的程度。平均增长速度不能由环比增长速度直接求出,而是要依据平均发展速度与平均增长速度之间的关系来推算,即

$$\text{平均增长速度} = \text{平均发展速度} - 1$$

## 任务四 了解长期趋势分析和季节变动分析

### 任务卡 1

请你写出你所了解的长期趋势分析的方法。

☐ 移动平均法

☐ \_\_\_\_\_

☐ 平均法

☐ \_\_\_\_\_

☐ \_\_\_\_\_

☐ \_\_\_\_\_





## 知识卡 1 长期趋势分析

长期趋势（用  $T$  表示）是时间序列的基本形式，是指现象在较长一段时期内，由于普通的、持续的、决定的基本因素的作用，使发展水平所呈现的逐渐增加向上发展或逐渐减少向下发展的趋势。例如，由于科学技术的日益发展，劳动生产率的不断提高，我国的国内生产总值与人均收入呈逐年提高的趋势。长期趋势经过若干年以后，也可能改变其变动方向，由上升趋势转变为下降趋势，或由下降趋势变为上升趋势。例如，在产品寿命周期中，处于成长期和成熟期的产品，其产量和利润均呈上升趋势，而成本则呈下降趋势；到了衰退期和更替期，由于新产品的出现，原有产品产量和利润转为下降趋势，而成本则转为上升趋势。

长期趋势分析就是运用一定的数学方法，对原序列进行加工整理，以排除季节变动、循环波动和不规则变动等因素的影响，显示出现象发展变化的长期趋势或规律，为预测和决策等管理活动提供依据的统计分析方法。长期趋势的测定方法很多，下面只介绍常用的移动平均法和趋势线配合法。

### 一、移动平均法

移动平均法是从时间序列的第一项指标数值开始，按照一定的时间间隔（即固定的项数），逐项移动求其序时平均数的修匀方法。这种方法考虑到了现象动态发展的连续性，“移动平均”的方式把一段一段的指标差异抽象化，从而消除偶然因素的影响，这样通过对原序列中不规则变动的修匀，把隐藏在原序列中的变动规律较为明显地反映出来。

例如，表 6-13 是某商场 2012—2021 年商品销售额资料，通过移动平均可显示出其长期增长的趋势，如图 6-1 所示。

表 6-13 某商场 2012—2021 年商品销售额资料

单位：亿元

| 年份   | 销售额  | 三项移动平均 | 五项移动平均 | 四项移动平均 | 四项修正移动平均 |
|------|------|--------|--------|--------|----------|
| 2012 | 4.80 |        |        |        |          |
| 2013 | 5.33 | 5.63   |        | 6.07   |          |
| 2014 | 6.76 | 6.49   | 6.16   | 6.50   | 6.29     |

(续表)

| 年份   | 销售额  | 三项移动平均 | 五项移动平均 | 四项移动平均 | 四项修正移动平均 |
|------|------|--------|--------|--------|----------|
| 2015 | 7.38 | 6.89   | 6.60   | 6.92   | 6.71     |
| 2016 | 6.54 | 6.97   | 7.04   | 7.11   | 7.02     |
| 2017 | 7.00 | 7.02   | 7.52   | 7.55   | 7.33     |
| 2018 | 7.52 | 7.89   | 7.84   | 8.16   | 7.86     |
| 2019 | 9.14 | 8.55   | 8.40   | 8.75   | 8.64     |
| 2020 | 8.98 | 9.35   |        |        |          |
| 2021 | 9.35 |        |        |        |          |

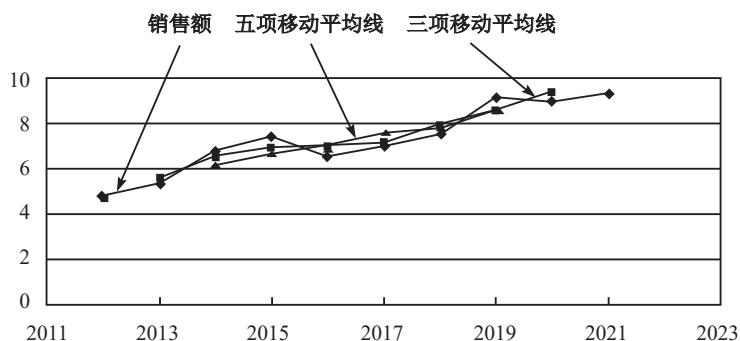


图6-1 移动平均法

由图 6-1 可以清楚地看出，移动平均线比原时间序列（销售额）的波动要小、要平滑，所以可通过移动平均线来反映销售额的长期趋势。

在移动平均法中，需要解决两个关键问题：一是移动平均的项数的确定，二是移动平均数放置的位置。

从图 6-1 可以看到，移动平均的项数越多，修匀的作用越大，移动平均线越平滑；但得到的移动平均数时间序列的数据较少，变动比较缓慢。移动平均的项数越少，修匀的作用越小，曲线的平滑度越差；但对原时间序列的变动反应比较敏感。因此，在确定移动平均的项数时，必须结合研究的目的，充分考虑到项数对移动平均结果的影响。一般情况下，往往是选择多个移动平均项数，从不同的角度刻画长期趋势。例如，在股市分析中，往往选择 5 天、10 天和 30 天移动平均（分别称为短期移动平均、中期移动平均和长期移动平均）各从不同的角度描述股价的变动趋势。

在移动平均法中，平均数放置的位置有两种方式。一种是把移动平均数放



置在平均时段的中间位置上。例如，在表 6-14 中，三项移动平均的第一个平均数 =  $\frac{4.80+5.33+6.76}{3}=5.63$ ，对正第二项的原值，以此类推；五项移动平均的第一个平均数 =  $\frac{4.80+5.33+6.76+7.38+6.54}{5}=6.16$ ，对正第三项的原值，以此类推。另一种是把平均数放置在平均时段的最后一个位置上。

表 6-14 移动平均法

| 年 份  | 销售额 / 亿元 | 三项移动平均 | 五项移动平均 |
|------|----------|--------|--------|
| 2012 | 4.80     |        |        |
| 2013 | 5.33     |        |        |
| 2014 | 6.76     | 5.63   |        |
| 2015 | 7.38     | 6.49   |        |
| 2016 | 6.54     | 6.89   | 6.16   |
| 2017 | 7.00     | 6.97   | 6.60   |
| 2018 | 7.52     | 7.02   | 7.04   |
| 2019 | 9.14     | 7.89   | 7.52   |
| 2020 | 8.98     | 8.55   | 7.84   |
| 2021 | 9.35     | 9.35   | 8.40   |

这两种方式的区别可通过图 6-2 来观察。

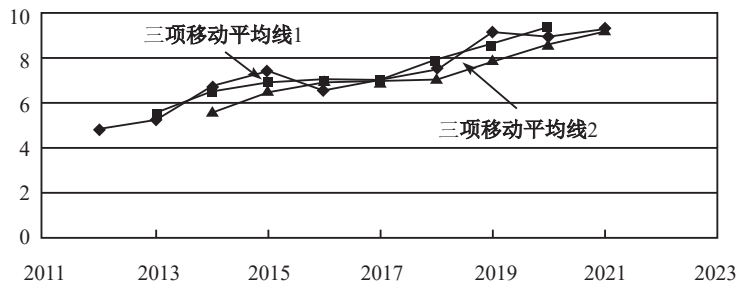


图 6-2 移动平均的两种方式

图 6-2 是对表 6-14 中的销售额按两种方式进行移动平均，其中三项移动平均线 1 是把平均数放置在平均时段的中间位置上，而三项移动平均线 2 是把平均数放置在平均时段的最后位置上。从图中可以看出，前者与原时间序列的拟合度很好，但存在时间上的“滞后”；后者能与原序列在时间上保持同步，但拟



合度稍差。研究者可根据这一特点,结合自己的研究目的选择移动平均的方式。Office 系统和股市分析中的移动平均法都采用第二种方式。

在采用第一种方式进行移动平均的时候要注意,如果移动平均的项数为偶数,必须进行两次移动平均。第一次移动平均的目的在于修匀数列,但位置无法与原序列一致,所以必须对第一次移动平均的结果再进行一次二项移动平均,以修正移动平均数的位置。例如,根据表 6-14 中的数据可以计算出,四项移动平均的第一个平均数  $= (4.80+5.33+6.76+7.38) \div 4 \approx 6.07$ ,对正第二项和第三项的原值,以此类推。这样会使移动平均数时间序列和原时间序列错过半期,无法直接进行比较,还必须进行一次修正平均,即再进行一次两项移动平均,这样新的平均序列的各期数值才能和原时间序列对准。当然,如果采取第二种移动平均的方式(即把平均数放置在平均时段的最后一个位置上),则不论是偶数项还是奇数项移动平均都可一次移动完成。

移动平均法也可借助于 Excel 来计算(此时,移动平均数放置在平均时段的最后一个位置上)。



想一想

## 二、趋势线配合法

趋势线配合法是在对原时间序列资料进行初步分析的基础上,根据其发展变化趋势的类型,用数学的方法配合适当的方程式,以反映现象长期变动趋势的一种方法。根据现象发展变化的趋势不同,趋势线配合法一般可分为直线趋势和曲线趋势两种。本书只讨论直线趋势方程的建立。

在对时间序列进行分析时,如果时间序列的数据大体上按逐期等量增加或减少,则可以认为这种现象的基本发展趋势是直线型,因而应配合相应的直线方程来反映其长期趋势。其直线方程为

$$y_t = a + bx$$

式中, $x$  为时间序列中的时间顺序值,是自变量; $y_t$  为时间序列的预测趋势值,是关于时间刻度  $x$  的函数; $a$  与  $b$  是两个待定系数。 $a$  为起始值,即当  $x=0$  时趋势直线在  $y$  轴上的截距。 $b$  为斜率,代表  $x$  每变动一个单位时间发展水平  $y$  的平均增减量。

如果方程中的两个待定系数  $a$  与  $b$  确定了,那么所配合的直线方程就唯一确定了。趋势线配合法是根据时间序列的数据资料来求解待定的系数,常用的



方法有两种：平均法和最小平方法。

### 1. 平均法

平均法是将全部时间序列资料分为相等的两部分（奇数项时舍弃最初一项资料），分别计算出各部分的平均数（ $\bar{x}$ ， $\bar{y}$ ）代入配合直线方程 $y_t = a + bx$ 中求解 $a$ 与 $b$ 的方法。利用平均法求配合直线的数学依据是：时间序列的实际值 $y$ 与相应的趋势值 $y_t$ 的离差之和等于0，即

$$\sum(y - y_t) = 0$$

现以 2016—2021 年某地区生产总值为例说明其计算过程，如表 6-15 所示。

表 6-15 平均法配合直线方程计算表

| 年 份  | 序 号 | 生产总值 $y_t$ / 万元 | 趋势值 $y_t$ | 离差 ( $y - y_t$ ) |
|------|-----|-----------------|-----------|------------------|
| 2016 | 1   | 95 933          | 81 317    | +14 616          |
| 2017 | 2   | 102 398         | 105 008   | -2 610           |
| 2018 | 3   | 116 694         | 128 700   | -12 006          |
| 合计   | 6   | 315 025         | 315 025   | 0                |
| 2019 | 4   | 136 515         | 152 390   | -15 875          |
| 2020 | 5   | 182 321         | 176 081   | +6 240           |
| 2021 | 6   | 209 407         | 199 772   | +9 635           |
| 合计   | 15  | 528 243         | 528 243   | 0                |

表中前一部分（2016—2018 年）：

$$\bar{x}_1 = \frac{1+2+3}{3} = 2$$

$$\bar{y}_1 = \frac{95\,933 + 102\,398 + 116\,694}{3} \approx 105\,008$$

表中后一部分（2019—2021 年）：

$$\bar{x}_2 = \frac{4+5+6}{3} = 5$$

$$\bar{y}_2 = \frac{136\,515 + 182\,321 + 209\,407}{3} = 176\,081$$

将上边两组平均数（ $\bar{x}_1$ ， $\bar{y}_1$ ），（ $\bar{x}_2$ ， $\bar{y}_2$ ）代入趋势直线方程得方程组：



$$105\,008=a+2b$$

$$176\,081=a+5b$$

解方程组得： $a=57\,626$ ， $b=23\,691$ 。

故所求直线方程为

$$y_t=57\,626+23\,691x$$

如果要预测 2025 年的国内生产总值，可将 2025 年的时间序号  $x$  取 10，则

$$y_{2025}=57\,626+23\,691 \times 10=294\,536 \text{ (亿元)}$$

利用平均法配合直线方程，虽然各年实际水平与趋势值的离差之和等于 0，离差的绝对值并不是最小值，因此这种方法配合的并不是最适应的直线。然而，平均法比较简便，在现象变动比较均衡时，用平均法配合趋势直线可以粗略地反映长期趋势。

## 2. 最小平方法

最小平方法又称最小二乘法，是建立趋势方程、分析长期趋势较为常用的方法。它是依据时间序列的观察值与趋势值的离差平方和为最小值的基本原理，拟合一种趋势模型，然后利用数学中求极值的方法来确定方程中的待定系数、建立方程。

用最小平方法拟合出来的趋势直线将比用平均法配合的趋势直线更为理想、合理，原因是用最小平方法建立趋势方程必须满足以下两个条件：

(1) 原时间序列中各期的指标数值 ( $y$ ) 与其对应的趋势值 ( $y_t$ ) 的离差平方和为最小值，即

$$\sum (y - y_t)^2 = \text{最小值}$$

(2) 原时间序列中各期的指标数值 ( $y$ ) 与其对应的趋势值 ( $y_t$ ) 的离差和等于零，即

$$\sum (y - y_t) = 0$$

设拟合的趋势直线方程为

$$y_t=a+bx$$

按照最小平方法的要求，则



$$D = \sum (y - y_i)^2 = \sum [y - (a + bx)]^2 = \sum (y - a - bx)^2 = \text{最小值}$$

将  $D$  看作关于  $a$  和  $b$  的函数, 要使  $D$  为最小值, 必须满足

$$\frac{\partial D}{\partial a} = \frac{\partial D}{\partial b} = 0$$

于是得

$$\frac{\partial D}{\partial a} = \sum 2(y - a - bx)(-1) = 0$$

$$\frac{\partial D}{\partial b} = \sum 2(y - a - bx)(-x) = 0$$

整理后得下列联立方程式:

$$\begin{cases} \sum y = na + b \sum x \\ \sum xy = a \sum x + b \sum x^2 \end{cases}$$

求解联立方程式, 得

$$\begin{cases} a = \frac{\sum y}{n} - b \frac{\sum x}{n} = \bar{y} - b\bar{x} \\ b = \frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - (\sum x)^2} \end{cases}$$

将时间序列中的时间  $x$  及对应的发展水平  $y$  一并代入  $a$ 、 $b$  公式, 即可求得系数  $a$  和  $b$ , 从而得到直线趋势方程  $y_t = a + bx$ 。实际计算的时候, 是用时间刻度来作为自变量  $x$  的。

【例 6-7】已知某企业 2013—2021 年产品销售额如表 6-16 所示。

表 6-16 某企业 2013—2021 年产品销售额

| 年 份      | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|----------|------|------|------|------|------|------|------|------|------|
| 销售额 / 万元 | 100  | 119  | 125  | 135  | 147  | 159  | 167  | 179  | 195  |

试用最小平方法配合趋势直线方程, 并预测 2022 年和 2023 年的销售额。

解: 用  $x=0, 1, \dots, 8$  分别代表时间 2013 年, 2014 年,  $\dots$ , 2021 年, 将计算过程表达在表 6-17 中。





表 6-17 最小平方法计算表

| 年 份  | 时间序号 $x$ | 销售额 $y$ / 万元 | $x^2$ | $xy$  |
|------|----------|--------------|-------|-------|
| 2013 | 0        | 100          | 0     | 0     |
| 2014 | 1        | 119          | 1     | 119   |
| 2015 | 2        | 125          | 4     | 250   |
| 2016 | 3        | 135          | 9     | 405   |
| 2017 | 4        | 147          | 16    | 588   |
| 2018 | 5        | 159          | 25    | 795   |
| 2019 | 6        | 167          | 36    | 1 002 |
| 2020 | 7        | 179          | 49    | 1 253 |
| 2021 | 8        | 195          | 64    | 1 560 |
| 合计   | 36       | 1 326        | 204   | 5 972 |

将表 6-17 中有关数值代入计算公式：

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{9 \times 5\,972 - 36 \times 1\,326}{9 \times 204 - 36^2} = \frac{53\,748 - 47\,736}{1\,836 - 1\,296}$$

$$= \frac{6\,012}{540} \approx 11.13$$

$$a = \frac{\sum y}{n} - b \frac{\sum x}{n} = \frac{1\,326}{9} - 11.13 \times \frac{36}{9}$$

$$\approx 147.3 - 44.5 = 102.8$$

即趋势直线方程为

$$y_t = 102.8 + 11.13x$$

将 2022 年的时间刻度  $x=10$  代入趋势方程得 2022 年销售额的预测值：

$$y_{2022} = 102.8 + 11.13 \times 10 = 214.1 \text{ (万元)}$$

同理，2023 年的销售额预测值为

$$y_{2023} = 102.8 + 11.13 \times 11 = 225.23 \text{ (万元)}$$

上述方法是最小平方法的一般方法，为了减少计算工作量，还可以采用以下简捷法。



当时间项数  $n$  为奇数时, 把正中间一期的时间刻度定义为 0,  $x$  的间隔期取 1, 即各期时间刻度为

$$-\frac{n}{2}, \dots, -3, -2, -1, 0, 1, 2, 3, \dots, \frac{n}{2}$$

当  $n$  为偶数时, 把正中间两期的时间刻度分别定义为 -1 和 1,  $x$  的间隔期取 2, 即各期时间刻度为

$$\dots, -5, -3, -1, 1, 3, 5, \dots$$

此时  $\sum x = 0$ , 于是  $a$  和  $b$  的计算公式可简化为

$$a = \frac{\sum y}{n} = \bar{y}$$

$$b = \frac{\sum xy}{\sum x^2}$$

计算过程得到大幅度简化。

由此可见, 简捷法和一般方法的预测结果完全一致。当然, 由于两种方法的原点不同, 直线方程的截距  $a$  是不同的。

### 任务卡 2

生活中有哪些商品的销售受季节影响? 举例说明。

---

---

## 知识卡 2 季节变动分析

季节变动 (S) 是指时间序列中的指标数值由于受各种因素的影响, 在一年内随着季节的更替而产生的周期性变动。例如, 夏季是冰激凌、电风扇、空调等产品的销售旺季, 而冬季是销售淡季, 销售量呈现出周期性的变动。

测定季节变动的方法主要有两种: 按月 (季) 平均法和趋势剔除法。



## 一、按月（季）平均法

按月（季）平均法是指不考虑现象长期趋势的影响，以历年的各月（季）平均数同全时期月（季）平均数相比求得季节指数来反映现象季节规律的方法。这种方法适合于不含长期趋势的季节变动分析。其计算过程：先分别计算各年同月（季）平均数，再计算各年所有月（季）的总平均数，然后将各年同月（季）平均数与总平均数进行对比，求得季节指数（可称为季节比率）。如果某月（季）的季节指数大于100%，则该月（季）为旺季；如果小于100%，则为淡季。季节指数的计算公式为

$$\text{季节指数 } a_i (\%) = \frac{\text{各年同月（季）平均数}}{\text{总的月（季）平均数}} \times 100\%$$

## 二、趋势剔除法

趋势剔除法是利用移动平均法来剔除长期趋势影响后再测定其季节指数的方法，这种方法适用于时间序列中包含长期趋势的季节变动分析。



素质园地



## 职业基础知识测试

### 一、填空题

1. 时间序列根据其指标表现形式可分为\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_。
2. 发展水平是指时间序列中每一项具体的\_\_\_\_\_。它具体反映社会经济现象在\_\_\_\_\_上所达到的规模和水平，通过不同时期发展水平的比较，可以给人\_\_\_\_\_的印象。
3. 由于社会经济现象发展的总速度不等于各年发展速度之积，而等于\_\_\_\_\_的连乘积，所以平均发展速度不能用算术平均法计算，而要用几何平均法，这种方法称为\_\_\_\_\_或\_\_\_\_\_。
4. 平均发展速度是各时间\_\_\_\_\_的序时平均数，它说明社会经济现象在较长一段时间中各期\_\_\_\_\_的程度。
5. 季节变动是指时间序列中的指标数值由于受各种因素的影响，在一年内随着\_\_\_\_\_而产生的周期性变动。



## 二、简答题

1. 简述时间序列的含义和作用。
2. 编制时间序列应遵守哪些原则？
3. 什么是增长量？什么是平均增长量？
4. 什么是发展速度和增长速度？
5. 什么是长期趋势分析？



## 职业核心能力训练

### 【实训目标】

- (1) 锻炼数据积累的意识，培养建立数据积累制度的观念。
- (2) 学会将动态分析方法运用于社会实践中。

### 【具体要求】

(1) 对于企业的相关数据，应恪守职业道德，予以保密，坚决杜绝用于其他用途。

(2) 调查本省（直辖市、自治区、特别行政区）三年内各月的旅游支出资料；或选择一个企业，了解其三年内各月的销售量（额）资料。用季节变动分析方法分析旅游支出或企业销售量（额）的季节变动规律。

### 【实训记录】

### 【交流与评价】

| 项 目  | 成 绩               |                   |                |     |
|------|-------------------|-------------------|----------------|-----|
|      | 收集资料的完整性<br>(30%) | 分析报告的科学性<br>(30%) | 结果正确性<br>(40%) | 总 分 |
| 自我评价 |                   |                   |                |     |
| 同学评价 |                   |                   |                |     |
| 教师评价 |                   |                   |                |     |



## 项目七 了解统计指数

### 知识目标 →

- 了解统计指数的含义和作用，理解统计指数的分类、综合指数的含义
- 掌握数量指标综合指数、质量指标综合指数的编制方法
- 掌握指数体系的含义和作用，掌握综合指标体系的两因素分析方法
- 掌握加权算术平均指标的含义和编制方法

### 能力目标 →

- 能够编制数量指标综合指数、质量指标综合指数
- 掌握加权算术平均指标的编制方法

### 思维导图 →

#### 了解统计指数

##### 认识统计指数

统计指数的含义

统计指数的种类和作用

##### 掌握综合指数

综合指数的含义和编制原理

数量指标综合指数和质量指标综合指数

##### 了解指数体系

指数体系的含义和作用

指数体系的两因素分析

##### 掌握平均指数

平均指数的含义及加权算术平均指数

加权调和平均指数



## 项目情境 →

道琼斯指数是世界上历史最为悠久的股票指数，它的全称为道琼斯工业平均指数。它是在 1884 年由道琼斯公司的创始人查尔斯·道开始编制的。最初的道琼斯指数根据 11 种具有代表性的铁路公司的股票，采用算术平均法进行计算编制而成，发表在查尔斯·道自己编辑的《每日通讯》上。其计算公式为：股票价格平均数 = 入选股票的价格之和 / 入选股票的数量。

道琼斯指数最初用简单算术平均法求得，当遇到股票的除权除息时，股票指数将发生不连续的现象。1928 年后，道琼斯指数改用新的计算方法，即在计点的股票除权或除息时采用连接技术，以保证股票指数的连续，从而使股票指数得到了完善，并逐渐推广到全世界。

通过上述内容你对统计指数有所了解了吗？道琼斯指数和统计指数有什么关系呢？

---

---



## 任务一 认识统计指数

### 任务卡 1

小张的父母特别爱炒股，所以他每天都会听到上证指数、股票指数等词语，小张很是疑惑，你如何向他解释清楚呢？

### 知识卡 1 统计指数的含义

指数的概念最初是从对物价变动的研究中产生的，并有一个逐渐发展的过程。从 1650 年英国人沃汉首创物价指数，迄今已有 300 多年历史。那时的物价指数只限于观察单个商品的价格变动，也就是现在的个体价格指数。之后，随着指数的应用范围不断扩大，其含义和内容也发生了很大变化：由单纯反映一种现象的相对变动，到反映多种现象的综合变动；由单纯的不同时间的对比分析，到不同空间的比较分析；等等。

指数的概念有广义和狭义之分。广义的指数是用以测定某个变量在时间或空间上变动程度和方向的相对数。它除了包括狭义的指数外，还包括计划完成程度相对数、动态相对数等。狭义的指数是一种特殊的相对数，它是说明由许多不能直接加总的要素所组成的复杂社会经济现象综合变动的相对数，如零售物价指数、消费价格指数、股价指数。

在社会经济现象中，很多时候只能考察个别现象的变动，如在研究商品销售情况时，只能就个别商品计算其发展速度来反映其动态规律。但要从整体上





综合考察多种商品销售情况的变动时,由于无法计算销售量的综合值,因此不能用发展速度指标刻画其综合变动,如表 7-1 的资料。

表 7-1 某副食品商店销售三种主要商品资料

| 商 品 | 计量单位 | 销 售 量    |           | 销售价格 / 元 |           |
|-----|------|----------|-----------|----------|-----------|
|     |      | 基期 $q_0$ | 报告期 $q_1$ | 基期 $p_0$ | 报告期 $p_1$ |
| 猪肉  | 千克   | 2 500    | 3 000     | 16       | 24        |
| 面粉  | 千克   | 17 400   | 18 600    | 1.8      | 2.4       |
| 洗衣粉 | 包    | 1 200    | 1 100     | 4.5      | 4         |

根据表 7-1 的资料,利用一般计算动态相对数的方法,只能分别计算每种商品销售量和价格的发展速度  $K_q$  和  $K_p$  :

$$K_{q_{\text{猪肉}}} = \frac{q_1}{q_0} = \frac{3\,000}{2\,500} = 120\%$$

$$K_{q_{\text{面粉}}} = \frac{q_1}{q_0} = \frac{18\,600}{17\,400} \approx 106.9\%$$

$$K_{q_{\text{洗衣粉}}} = \frac{q_1}{q_0} = \frac{1\,100}{1\,200} \approx 91.7\%$$

$$K_{p_{\text{猪肉}}} = \frac{p_1}{p_0} = \frac{24}{16} = 150\%$$

$$K_{p_{\text{面粉}}} = \frac{p_1}{p_0} = \frac{2.4}{1.8} \approx 133.3\%$$

$$K_{p_{\text{洗衣粉}}} = \frac{p_1}{p_0} = \frac{4.0}{4.5} \approx 88.9\%$$

但要刻画三种商品销售量和价格的综合变动,发展速度就无能为力了。

再如,工业产品的产量,我们可以就个别种类,如电视机、汽车、钢材等用速度指标分别测定其变动程度,但要把品种繁多的工业产品作为一个总体考察,以反映全部产品产量的总变动,则用速度指标不能直接测定。之所以如此,是因为不同的商品或产品具有不同的使用价值,不同价格水平又分属于不同的实物产品,它们之间直接相加不具有任何经济意义,这在统计上称为不能同度量。然而,实际工作中又需要从整体上反映这种不能同度量现象的综合变动,



这就产生了解决这一问题的方法——狭义指数。下面从狭义的角度讨论指数的编制方法及其应用。

### 任务卡 2

请根据自己的理解，勾选出总指数。

- |                                  |                                  |
|----------------------------------|----------------------------------|
| <input type="checkbox"/> 个体指数    | <input type="checkbox"/> 工业总产量指数 |
| <input type="checkbox"/> 零售物价总指数 | <input type="checkbox"/> 单项指数    |
| <input type="checkbox"/> 环比指数    | <input type="checkbox"/> 商品销售量指数 |
- 
- 

## 知识卡 2 统计指数的种类和作用

### 一、统计指数的种类

根据不同研究目的，可对指数进行不同的分类。

#### 1. 按反映的对象范围分类

按反映的对象范围不同，统计指数可分为个体指数和总指数。

个体指数是反映个体现象变动的相对数，又称单项指数，实际上就是个体现象的发展速度。例如，表 7-1 中三种商品各自的销售量发展速度和价格发展速度，就是其销售量个体指数和价格个体指数。总指数是反映经济现象动态变动的相对数，亦即狭义的指数，如零售物价总指数、商品销售量指数、工业总产量指数等。此外，在个体指数和总指数之间还存在类指数，它是说明现象总体中各类现象总变动的指数。例如，在零售商品价格总变动中的食品类价格指数、衣着类价格指数等。总指数与类指数之间的划分是相对的，没有绝对界限。类指数对总指数而言具有个体指数的性质，对个体指数而言又具有总指数的性质。

#### 2. 按指数化指标性质分类

按指数化指标性质不同，统计指数可分为数量指标指数和质量指标指数。



数量指标指数是根据数量指标计算的,反映社会经济现象总的规模和水平的变动,如产量指数、销售量指数、职工人数指数等。质量指标指数是根据质量指标计算的,反映现象内涵数量关系变化的指数,如价格指数、劳动生产率指数、单位成本指数等。应当指出,这种划分具有一定的相对性。有些指数,在某种情况下是数量指标指数,在另一种情况下可以成为质量指标指数;在某种情况下是质量指标指数,在另一种情况下也可以成为数量指标指数。

### 3. 按对比指标形式分类

按对比指标形式不同,统计指数可分为总量指标指数和平均指标指数。

总量指标指数是利用总量指标的对比来反映现象总量变动的指数,如商品销售量指数、产品产量指数等。平均指标指数是利用平均指标的对比以反映平均指标变动的指数,如劳动生产率指数、平均工资指数等。

### 4. 按其采用的基期形式分类

按其采用的基期形式不同,统计指数可分为定基指数和环比指数。

定基指数是指在一个指数数列中,按照某一固定基期所编制的指数,它反映某种社会经济现象长期的变动程度,如我国以 2021 年为固定基期计算的国内生产总值指数、工业总产值指数和农业总产值指数等。环比指数是指在一个指数数列中,各时期的指数以其前一时期为基期所编制的指数。它反映某种社会经济现象逐期的变动程度,如按月、季、年连续计算的产量指数、价格指数或成本指数等。

在指数方法中,计算个体指数是比较简单的。因此,指数理论主要研究不能同度量现象综合变动的总指数。编制总指数是指数理论的核心,其编制方法主要有综合指数和平均数指数两种,我们将在本项目分别讨论这两种方法。

## 二、统计指数的作用

指数在社会经济统计工作中应用广泛,主要体现在以下两个方面:

(1) 综合反映复杂现象的总体数量变动的方向、程度和绝对效果,这是指数最基本的作用。编制指数可以将由许多不能同度量的个别事物组成的现象总体过渡到可以加总对比的状态,从而综合说明现象总体变动的程度。同时,子项和母项的差额可以说明现象变动的绝对效果。

(2) 对现象数量总变动进行因素分析。许多现象的总变动都是其内部诸多



因素综合影响的结果。如商品销售额受销售量和销售价格两个因素的影响,工业总产值受职工人数、劳动生产率和产品价格三个因素的影响,劳动生产率受各类人员劳动生产率水平和人员构成的影响等。通过编制指数并借助于指数体系,可以分析和测定复杂社会经济现象变动受各构成因素变动的影响。



素质园地

## 任务二 掌握综合指数

### 任务卡 1

你了解综合指数吗?请查找资料,写下你的收获。

---

---

### 知识卡 1 综合指数的含义和编制原理

#### 一、综合指数的含义

综合指数是编制总指数的基本形式,它是按照“先综合,后对比”的思路编制的总指数。

总指数要反映不能同度量现象的总变动,必须将不能同度量的现象转化为能同度量的现象,然后综合得出总量指标并通过对比编制出总指数。例如,要研究多种工业产品产量的总变动,首先碰到的一个问题是各种不同类的产品产量不能直接相加,如一万吨钢、一万台电视和一万米棉布不能直接相加。因为不同种类产品的使用价值不同、计量单位不同,加在一起没有意义。但是,在商品货币存在的条件下,各种商品有一个共同点,即它们都是没有差别的人类劳动的一般凝结物,这就是商品的价值量。商品的价值量都是同质的,是相



加的。因此，综合指数的编制，首先要将不能同度量的现象转化为能同度量的价值形态。

如何实现这种转化，这就要求寻找一个能使不同度量的现象过渡到能够同度量的现象的因素，即同度量因素。同度量因素是将所要研究的因素——指数化因素过渡到能够同度量的媒介和桥梁。例如，将不同计量单位的产品或商品的数量乘其价格，均可过渡到价值量形态——商品销售额或总产值：

$$\text{商品销售量} \times \text{价格} = \text{商品销售额}$$

$$\text{产品产量} \times \text{价格} = \text{总产值}$$

这样，具有不同计量单位的产品产量和商品销售量通过价格这个因素过渡到了可以同度量的价值量。在这里，价格这个因素就是同度量因素。在不同情况下，同度量因素是不相同的。如要研究各种商品价格的变动，则可以用销售量作为同度量因素等。

根据以上分析，可以概括出编制综合指数的一般原则：通过引入同度量因素，把不能直接加总的因素转化为同度量的价值量指标，然后通过价值量指标的对比来反映因素的综合变动。在编制数量指数的时候要以质量指标为同度量因素，在编制质量指数的时候要以数量指标为同度量因素。

同度量因素虽然解决了指数化因素不能直接加总的问题，但是，通过转化综合而来的价值总量的变动中既包含指数化因素的变动，也包含同度量因素的变动。为了刻画指数化因素的综合变动，就必须剔除价值总量中同度量因素变动的影响，即在编制综合指数时，必须采用同一时期的指标作为同度量因素，这样才能纯粹研究指数化因素的变动。

为了测定指数化因素的总变动，我们采用同一时期的同度量因素。这个同度量因素既可以是报告期的，也可以是基期的，但是，究竟应该选择哪个时期的，也是编制综合指数的重要问题。因为采用的时期不同，会得到不同的计算结果，具有不同的经济内容。

## 二、综合指数的编制原理

要编制总指数，用一个数值反映出多种事物的综合动态，首先必须解决不同事物数量不同度量的问题，设法将不能加总变为可以加总，然后才能进行对比。虽然不同产品有不同的价格、产量和单位成本，但是它们的价值量，如总



产值、销售额、总成本等都具有相同的计量单位，故可直接进行加总。因此，在计算某个因素指标的变动程度时，加入另一个因素指标，用来转化为同度量。例如，在编制价格指数时，可以加入产量指标，用各种产品的价格乘以各自的产量，就可以将不同度量的价格转化为同度量的总产值。在编制总指数的过程中，加入的能够将不同度量的经济指标转化为同度量经济指标的媒介指标，称为同度量因素。加入同度量因素，将不同度量的指标转化为同度量的指标后，就可以把各种产品或商品的价值量加总起来，再将两个时期的价值量进行对比来编制指数。有了同度量因素，就可以计算价格总指数、产量总指数及单位成本指数，它们的计算公式如下：

以产量或销售量  $Q$  为同度量因素时，价格总指数为

$$\bar{K}_p = \frac{\sum P_1 Q}{\sum P_0 Q}$$

以价格  $P$  为同度量因素时，产量总指数为

$$\bar{K}_Q = \frac{\sum P Q_1}{\sum P Q_0}$$

以产量  $M$  为同度量因素时，单位成本总指数为

$$\bar{K}_z = \frac{\sum z_1 M}{\sum z_0 M}$$

如何使两个时期的价值量之比单纯反映指数化指标的变动程度呢？以产量指数为例，既然总产值是产量和价格两个因素指标的乘积，那么，如果报告期产值与基期产值均按同一时期的价格计算，即价格不变，则对比结果当然只反映产量的变动程度。同理，在计算价格总指数时，只要报告期与基期的产值均按同一时期的产量计算，即在对比时就剔除了产量变动，则计算结果就单纯反映价格的变动程度。

综上所述，综合指数的基本编制原理可以表述为：首先，在分析复杂社会经济现象的综合变动时，借助同度量因素进行综合对比；其次，在编制指数时，固定同度量因素来测定人们所关心的因素的变动；最后，将两个时期的总量指标进行对比，以测定指数化指标的数量变动程度。





### 任务卡 2

小袁周末逛商场时无意间看到了标签上面有商品销售量指数，于是向销售员询问。假如你是销售员，你如何向他解释？

## 知识卡 2 数量指标综合指数和质量指标综合指数

### 一、数量指标综合指数

数量指数是说明数量指标综合变动情况的总指数，如工业品产量指数、职工人数指数、商品销售量指数等。现以表 7-1 为例，讨论商品销售量指数的编制原理。

通过上述分析可以看出，要刻画三种商品销售量的综合变动（即编制三种商品的销售量指数），式（7-1）是不行的，因为销售量的计量单位不同，将销售量直接相加没有意义。

$$\frac{\sum q_1}{\sum q_0} = \frac{3\,000 + 18\,600 + 1\,100}{2\,500 + 17\,400 + 1\,200} \quad (7-1)$$

因此，要想将销售量相加，就必须把销售量转化为价值量指标，这时价格作为同度量因素必须介入，自然会想到下面的公式：

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{3\,000 \times 24 + 18\,600 \times 2.4 + 1\,100 \times 4}{2\,500 \times 16 + 17\,400 \times 1.8 + 1\,200 \times 4.5} \quad (7-2)$$

但在式（7-2）中，尽管把销售量转化为价值量指标（即销售额），解决了综合性问题，但刻画的内容却发生了变化。它所反映的是销售额的变动，里面既包含销售量的变动，又包含价格的变动。为了仅仅刻画销售量的变动，需要将式（7-2）中的价格因素固定起来保持不变。也就是说，分子和分母必须采用





同一时期的价格。当然,既可以选择基期的价格,也可以选择报告期的价格作为同度量因素,于是形成了以下两种编制销售量总指数的形式。

### 1. 以基期价格作为同度量因素

此时,销售量指数可用式(7-3)来计算:

$$L_q = \frac{\sum p_0 q_1}{\sum p_0 q_0} \quad (7-3)$$

这个公式是由德国人拉斯贝尔于1864年提出的,称为拉氏数量指数。其计算结果说明,在基期价格水平情况下,销售量的综合变动程度。分子分母的绝对差额  $\sum p_0 q_1 - \sum p_0 q_0$  说明由于商品销售量变动对销售额影响的绝对程度。以表7-1资料为例,其拉氏销售量指数为

$$L_q = \frac{\sum p_0 q_1}{\sum p_0 q_0} = \frac{3000 \times 16 + 18600 \times 1.8 + 1100 \times 4.5}{2500 \times 16 + 17400 \times 1.8 + 1200 \times 4.5} = \frac{86430}{76720} \approx 112.66\%$$

分子分母的绝对差额为

$$\sum p_0 q_1 - \sum p_0 q_0 = 86430 - 76720 = 9710 \text{ (元)}$$

计算结果表明,三种商品的销售量报告期比基期平均上升了12.66%,由于销售量的增加而使商品销售额增加了9710元。

### 2. 以报告期价格作为同度量因素

这时,销售量指数可用式(7-4)来计算:

$$P_q = \frac{\sum p_1 q_1}{\sum p_1 q_0} \quad (7-4)$$

这个公式是由德国人帕舍于1874年提出的,称为帕氏数量指数。其计算结果表明,在报告期价格水平条件下,销售量的综合变动程度。分子、分母的绝对差额  $\sum p_1 q_1 - \sum p_1 q_0$  表明由于销售量的变动使得销售额增长的绝对程度。

根据表7-1资料计算出帕氏指数为

$$P_q = \frac{\sum p_1 q_1}{\sum p_1 q_0} = \frac{3000 \times 24 + 18600 \times 2.4 + 1100 \times 4}{2500 \times 24 + 17400 \times 2.4 + 1200 \times 4} = \frac{121040}{106560} \approx 113.59\%$$



分子、分母的绝对差额为

$$\sum p_1 q_1 - \sum p_1 q_0 = 121\,040 - 106\,560 = 14\,480 \text{ (元)}$$

计算结果表明,商品销售量报告期比基期平均增加了 13.59%,由于商品销售量增加使商品销售额上升了 14 480 元。

根据以上例子可以看出,选用不同时期的价格作为同度量因素,编制的销售量综合指数是不同的。那么,究竟采用哪种公式计算销售量综合指数呢?这就必须根据其编制目的来选定。编制销售量指数的目的在于测定各种商品销售量的总变动,这就要求在计算中必须尽量排除价格变动的影响。因此,只有采用将销售价格固定在基期作同度量因素的公式,才能反映销售量本身的变动程度及影响的绝对销售额,而以报告期价格为同度量因素的公式,则包含了价格和销量的共同影响,不能确切反映销售量本身变动及其影响的绝对销售额。

上述对商品销售量综合指数的编制具有普遍意义。由此,我们可以概括出编制数量指标综合指数的一般原则:编制数量指标综合指数应将作为同度量因素的质量指标固定在基期,但也并不是绝对的。在实际应用中,要注意根据研究目的及资料条件等,针对具体情况灵活运用。如我国在编制工业产品物量指数时,采用某一固定时期的价格为同度量因素。其计算公式为

$$\bar{K}_q = \frac{\sum p_n q_1}{\sum p_n q_0}$$

式中,  $p_n$  为某一固定时期的价格。

## 二、质量指标综合指数

质量指数是反映质量指标综合变动情况的总指数,如价格指数、劳动生产率指数、单位成本指数等。现以表 7-1 资料为例,讨论质量指数的编制原理。

像销售量指数的编制原理一样,要刻画三种商品价格的综合变动,必须引入销售量作为同度量因素。但式(7-2)刻画的是销售额的变动,里面既包含价格的变动,又包含销售量的变动。为了仅反映价格的综合变动,就需要在式(7-2)中把销售量因素固定下来保持不变,即其分子、分母的销售量采用同一时期的。这样就有两种编制价格指数的方式,一是分子、分母都采用基期销售量为同度量因素,二是分子、分母都采用报告期销售量为同度量因素。



### 1. 以基期销售量为同度量因素

此时, 价格指数可用式(7-5)来计算:

$$L_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} \quad (7-5)$$

这个公式也是由拉斯贝尔提出的, 称为拉氏质量指数。其计算结果说明, 在基期销售量不变的情况下, 商品价格的综合变动程度。其分子、分母的绝对差额  $\sum p_1 q_0 - \sum p_0 q_0$  表明由于价格变动引起销售额变动的绝对程度。

仍以表 7-1 资料为例, 其拉氏价格指数为

$$L_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{24 \times 2\,500 + 2.4 \times 17\,400 + 4 \times 1\,200}{16 \times 2\,500 + 1.8 \times 17\,400 + 4.5 \times 1\,200} = \frac{106\,560}{76\,720} \approx 138.89\%$$

分子、分母的绝对差额为

$$\sum p_1 q_0 - \sum p_0 q_0 = 106\,560 - 76\,720 = 29\,840 \text{ (元)}$$

计算结果表明, 三种商品的价格总的来说报告期比基期平均上升了 38.89%。由于价格上涨使得销售额增加了 29 840 元。也说明居民在基期购买量条件下, 由于价格上涨使得报告期比基期多支付了 29 840 元。

### 2. 以报告期销售量为同度量因素

这时, 可用式(7-6)来计算价格指数:

$$P_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \quad (7-6)$$

这个公式由帕舍提出, 称为帕氏质量指数。其计算结果表明, 在报告期销售量条件下, 商品价格的综合变动程度。其分子、分母的绝对差额  $\sum p_1 q_1 - \sum p_0 q_1$  表明由于价格变动使得销售额变动的绝对程度。

以表 7-1 资料为例, 其帕氏价格指数为

$$P_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{24 \times 3\,000 + 2.4 \times 18\,600 + 4 \times 1\,100}{16 \times 3\,000 + 1.8 \times 18\,600 + 4.5 \times 1\,100} = \frac{121\,040}{86\,430} \approx 140.04\%$$

分子、分母的绝对差额为

$$\sum p_1 q_1 - \sum p_0 q_1 = 121\,040 - 86\,430 = 34\,610 \text{ (元)}$$



计算结果表明,该商店三种商品价格报告期比基期平均上升了 40.04%。由于价格上涨使得销售额增加了 34 610 元。它也说明居民在报告期购买量条件下,由于价格上涨而增加开支 34 610 元。

从以上例子可以看出,选用不同时期的销售量作为同度量因素,编制的价格指数也是不同的。用基期的销售量为同度量因素的公式,能够单纯反映商品价格的总变动,而计算表明的是居民在按过去的购买量条件下购买这三种商品支出的金额。而用报告期作为同度量因素的公式,尽管在反映商品销售价格的同时也包含有销售量的变动因素部分影响在内,但是,它可以说明在目前的商品销售量条件下,由于价格变动使商品销售额的变动情况,同时可以说明居民在目前的购买量条件下,由于物价变动而使支出额变动的差额,这样更具有现实意义。

以上介绍的商品价格综合指数的编制方法具有普遍意义,我们据此推广得出编制质量指标综合指数的一般原则:编制质量指标指数应将作为同度量因素的数量指标固定在报告期。

【例 7-1】设某粮油零售市场 2020 年和 2021 年三种商品的销售价格和销售量如表 7-2 所示。试分别以基期的价格和报告期的销售量为同度量因素,计算三种商品的销售量指数和价格综合指数,并加以分析。

表 7-2 某粮油零售市场 2020 年和 2021 年三种商品的销售价格和销售量

| 商 品 | 计 量 单 位 | 单 价 / 元 |        | 销 售 量  |        |
|-----|---------|---------|--------|--------|--------|
|     |         | 2020 年  | 2021 年 | 2020 年 | 2021 年 |
| 粳米  | 吨       | 2 600   | 3 000  | 120    | 150    |
| 标准粉 | 吨       | 2 300   | 2 100  | 150    | 200    |
| 花生油 | 千克      | 9.8     | 10.5   | 1 500  | 1 600  |

解:编制数量指标综合指数,要以基期的质量指标为同度量因素,即采取拉氏指数形式。为此,计算相关数据,如表 7-3 所示。



表 7-3 某粮油市场销售量指数计算表

| 商 品 | 计 量 单 位 | 销售量             |                 | 2020 年价格<br>$p_0$ / 元 | 2021 年销售<br>额 $p_0q_0$ / 元 | 假定性销售额<br>$p_0q_1$ / 元 |
|-----|---------|-----------------|-----------------|-----------------------|----------------------------|------------------------|
|     |         | 2020 年<br>$q_0$ | 2021 年<br>$q_1$ |                       |                            |                        |
| 粳米  | 吨       | 120             | 150             | 2 600                 | 312 000                    | 390 000                |
| 标准粉 | 吨       | 150             | 200             | 2 300                 | 345 000                    | 460 000                |
| 花生油 | 千克      | 1 500           | 1 600           | 9.8                   | 14 700                     | 15 680                 |
| 合计  |         |                 |                 |                       | 671 700                    | 865 680                |

则销售量指数为

$$L_q = \frac{\sum p_0 q_1}{\sum p_0 q_0} = \frac{865\,680}{671\,700} \approx 128.88\%$$

分子、分母的绝对差额为

$$\sum p_0 q_1 - \sum p_0 q_0 = 865\,680 - 671\,700 = 193\,980 \text{ (元)}$$

结果表明,与 2020 年相比,2021 年三种商品的销售量平均增加了 28.88%。销售量的增加使得销售额增加了 193 980 元。

编制质量指数,要以报告期的数量指标为同度量因素。为此计算出相关数据,如表 7-4 所示。

表 7-4 某粮油市场价格指数计算表

| 商 品 | 计 量 单 位 | 价格 / 元          |                 | 2021 年销售量<br>$q_1$ | 2021 年销售<br>额 $p_1q_1$ / 元 | 假定性销售额<br>$p_0q_1$ / 元 |
|-----|---------|-----------------|-----------------|--------------------|----------------------------|------------------------|
|     |         | 2020 年<br>$p_0$ | 2021 年<br>$p_1$ |                    |                            |                        |
| 粳米  | 吨       | 2 600           | 3 000           | 150                | 450 000                    | 390 000                |
| 标准粉 | 吨       | 2 300           | 2 100           | 200                | 420 000                    | 460 000                |
| 花生油 | 千克      | 9.8             | 10.5            | 1 600              | 16 800                     | 15 680                 |
| 合计  |         |                 |                 |                    | 886 800                    | 865 680                |

则价格指数为

$$P_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{886\,800}{865\,680} \approx 102.44\%$$

分子、分母的绝对差额为



$$\sum p_1 q_1 - \sum p_0 q_1 = 886\,800 - 865\,680 = 21\,120 \text{ (元)}$$

结果表明，与 2020 年相比，2021 年三种商品的价格平均上涨了 2.44%。价格上涨使得销售额增加了 21 120 元。

## 任务三 了解指数体系

### 任务卡 1

你了解指数体系吗？查找资料，写下你的收获。

### 知识卡 1 指数体系的含义和作用

#### 一、指数体系的含义

在经济分析中，一个指数通常只能说明某一方面的问题，而实践中往往需要将多个指数结合起来加以运用，这就要求建立相应的指数体系。

广义的指数体系类似于指标体系的概念，泛指由若干个内容上相互关联的统计指数所结成的体系。根据考察问题的需要构成这种体系的指数可多可少。

狭义的指数体系仅指几个指数在一定的经济基础上所结成的较为严密的数量关系式。其最为典型的表现形式是：一个总值指数等于若干个（两个或两个以上）因素指数的乘积。下面主要就狭义的指数体系进行讨论。例如：

$$\text{销售额} = \text{销售量} \times \text{销售价格}$$

这种静态数量关系也同样存在于指数之间，即

$$\text{销售额指数} = \text{销售量指数} \times \text{销售价格指数}$$



$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_0 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

此外,其绝对差额之间也存在一定的数量关系:

$$\sum p_1 q_1 - \sum p_0 q_0 = (\sum p_0 q_1 - \sum p_0 q_0) + (\sum p_1 q_1 - \sum p_0 q_1)$$

这说明销售额的变动额,恰好是销售量变动所引起的销售额变动额与销售价格变动所引起的销售额变动额共同组成的。

类似于销售额变动的情况有很多,例如:

$$\text{产品总成本} = \text{产品产量} \times \text{单位成本}$$

所以:

$$\text{产品总成本指数} = \text{产品产量指数} \times \text{单位成本指数}$$

又如:

$$\text{工业总产值} = \text{职工人数} \times \text{全员劳动生产率} \times \text{出厂价格}$$

所以:

$$\text{工业总产值指数} = \text{职工人数指数} \times \text{全员劳动生产率指数} \times \text{出厂价格指数}$$

$$\frac{\sum T_1 q_1 p_1}{\sum T_0 q_0 p_0} = \frac{\sum T_1 q_0 p_0}{\sum T_0 q_0 p_0} \cdot \frac{\sum T_1 q_1 p_0}{\sum T_1 q_0 p_0} \cdot \frac{\sum T_1 q_1 p_1}{\sum T_1 q_1 p_0}$$

这些都称为指数体系。这里主要讨论由两因素构成的指数体系,三因素指数体系(如上面的工业总产值指数体系)的分析思路与之类似。

## 二、指数体系的作用

### 1. 指数体系是因素分析的基本依据

编制指数不仅在于反映复杂社会经济现象的总变动,还要分析现象总变动中各构成因素的影响作用。利用指数体系可以对现象总变动中的影响因素进行定量分析,测定各因素变动对现象总变动在方向、程度和绝对量上的影响。这种分析方法又称指数因素分析法,它是统计分析中广泛运用的一种重要的分析方法。

例如,假定某商场本月的总销售额比上月增加了500万元,这种变动是由两个因素导致的:价格和销售量。如果计算出价格指数及其绝对差额以及销售量指数及其绝对差额,就可以清楚地看出价格和销售量对总销售额变动的影响,可以为商场有针对性地开展促销活动提供依据。





## 2. 指数体系可以进行统计推算

在统计研究或统计分析中,常常缺乏一些必要的统计资料。为此,需要按照社会经济现象的客观联系,根据已有的统计资料推算出所需要的统计资料。指数体系的这种经济和数量关系使我们可以根据现象之间的相互联系进行相互推算。

【例 7-2】某城市 2020 年工业企业平均职工人数为 120 万人,比上年增长 5%,全年工资总额为 11 211 200 万元,比上年增长 754 280 万元,要求推算 2019 年工人平均工资。

根据指数体系:

工资总额指数 = 平均职工人数指数 × 平均工资指数

可以先推算出工人年平均工资的变动:

工人年平均工资指数 = 工资总额指数 ÷ 平均职工人数指数

$$\begin{aligned} &= \frac{11\,211\,200}{11\,211\,200 - 754\,280} \div \frac{120}{120 \div (1 + 5\%)} \\ &= 1.072 \div 1.05 \\ &\approx 1.021 \\ &= 102.1\% \end{aligned}$$

计算结果表明,2020 年工人平均工资比 2019 年平均增长 2.1%。又知:

2020 年工人平均工资 =  $11\,211\,200 \div 120 = 93\,426.67$  (元)

由此,可以推算出 2019 年工人平均工资:

2019 年工人平均工资 =  $93\,426.67 \div 102.1\% = 91\,505.06$  (元)

### 任务卡 2

请根据自己的理解,勾选出指数体系的两因素分析。

☐ 两因素分析

☐ 指数因素分析法

☐ 多因素分析

☐ 总量指标变动的因素分析

☐ 平均指标变动的因素分析

☐ 总量因素分析法



## 知识卡 2 指数体系的两因素分析

社会经济现象数量的总变动是多因素综合作用的结果。指数在分析研究现象数量的总变动中各构成因素影响的方向、程度和绝对效果时,往往采用指数因素分析法。在指数因素分析法中,按照分析时所包含的因素多少不同分为两因素分析和多因素分析,按照分析现象的指数形式不同分为总量指标变动的因素分析和平均指标变动的因素分析。

这里主要研究总量指标变动的两因素分析。它主要有两种:综合指数体系的两因素分析和综合指数变形的平均指数体系的两因素分析。

### 一、综合指数体系的两因素分析

综合指数体系的两因素分析是将分析现象的总量指标分解为两个因素,分别测定其中每一个因素变动对总量指标总变动的影响。

综合指数体系的因素分析,依据的原理是如下两个公式:

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_0 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

$$\sum p_1 q_1 - \sum p_0 q_0 = (\sum p_1 q_1 - \sum p_0 q_1) + (\sum p_0 q_1 - \sum p_0 q_0)$$

如果把上述公式中的每一个部分都计算出来,就可以清晰地了解总量变动情况及各因素的影响情况。

【例 7-3】某蔬菜商场四种蔬菜的销售量和价格如表 7-5 所示。试借助指数体系从相对数和绝对数两方面分析销售额变动情况,以及销售量和价格对销售额的影响。

表 7-5 某蔬菜商场四种蔬菜的销售量和价格

| 蔬菜品种 | 销售量 / 千克 |          | 价格 / (元 · 千克 <sup>-1</sup> ) |          |
|------|----------|----------|------------------------------|----------|
|      | 上月 $q_0$ | 本月 $q_1$ | 上月 $p_0$                     | 本月 $p_1$ |
| 白菜   | 550      | 560      | 1.6                          | 1.8      |
| 黄瓜   | 224      | 250      | 2                            | 1.9      |
| 萝卜   | 308      | 320      | 1                            | 0.9      |
| 西红柿  | 168      | 170      | 2.4                          | 3        |



解：根据表 7-5 的资料可以计算出商场销售额的总变动情况。

本月销售额为

$$\begin{aligned}\sum p_1q_1 &= 560 \times 1.8 + 250 \times 1.9 + 320 \times 0.9 + 170 \times 3 \\ &= 2\,281 \text{ (元)}\end{aligned}$$

上月销售额为

$$\begin{aligned}\sum p_0q_0 &= 550 \times 1.6 + 224 \times 2 + 308 \times 1 + 168 \times 2.4 \\ &= 2\,039.2 \text{ (元)}\end{aligned}$$

则销售额指数为

$$K_{pq} = \frac{\sum p_1q_1}{\sum p_0q_0} = \frac{2\,281}{2\,039.2} \approx 111.86\%$$

销售额增长了

$$\sum p_1q_1 - \sum p_0q_0 = 2\,281 - 2\,039.2 = 241.8 \text{ (元)}$$

这说明，该商场四种蔬菜的销售额本月比上月增长了 11.86%，即增加了 241.8 元。

显然销售额的增加是由两个因素带来的：一是价格，二是销售量。它们分别使得销售额增加（或减少）了多少呢？为此，就要计算价格指数和销售量指数及绝对差额。

销售量指数一般采用拉氏指数：

$$L_q = \frac{\sum p_0q_1}{\sum p_0q_0} = \frac{2\,124}{2\,039.2} \approx 104.16\%$$

绝对差额为

$$\sum p_0q_1 - \sum p_0q_0 = 2\,124 - 2\,039.2 = 84.8 \text{ (元)}$$

结果表明：这四种蔬菜的销售量综合来看，本月比上月增长了 4.16%，从而使销售额增加了 84.8 元。

销售价格指数一般采用帕氏指数：

$$P_p = \frac{\sum p_1q_1}{\sum p_0q_1} = \frac{2\,281}{2\,124} \approx 107.39\%$$



绝对差额为

$$\sum p_1 q_1 - \sum p_0 q_1 = 2\,281 - 2\,124 = 157 \text{ (元)}$$

这表明这四种蔬菜的销售价格综合来看,本月比上月上漲了 7.39%,从而使销售额增加了 157 元。

销售额指数、销售量指数和销售价格指数三者形成以下指数体系:

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_0 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

$$11.86\% = 104.16\% \times 107.39\%$$

$$\begin{aligned} \sum p_1 q_1 - \sum p_0 q_0 &= (\sum p_1 q_1 - \sum p_0 q_1) + (\sum p_0 q_1 - \sum p_0 q_0) \\ 241.8 &= 84.8 + 157 \end{aligned}$$

据此综合分析说明:本月与上月相比,销售额增加了 11.86%,即增加了 241.8 元,其中由于销售量增加了 4.16% 使销售额增加 84.8 元,由于销售价格上涨了 7.39% 使销售额增加 157 元。由此可见,销售额增加的 241.8 元中,有接近 65% 的份额是价格上涨导致的,只有 35% 左右是由销售量增加导致的。

运用综合指数体系进行因素分析时要注意以下问题:

(1) 两个因素指数必须一个采用拉氏指数形式,另一个采用帕氏指数形式。通常数量指数采用拉氏指数,质量指数采用帕氏指数。当然,也可以数量指数采用帕氏指数,质量指数采用拉氏指数,这时指数体系依然成立:

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_1}{\sum p_1 q_0} \cdot \frac{\sum p_1 q_0}{\sum p_0 q_0}$$

$$\sum p_1 q_1 - \sum p_0 q_0 = (\sum p_1 q_1 - \sum p_1 q_0) + (\sum p_1 q_0 - \sum p_0 q_0)$$

(2) 指数体系因素分析法的前提条件是指数体系必须成立。有些现象的因素分析不能采用这种方法来进行,如导致销售额增加的因素有广告因素、促销因素、季节因素等,由于销售额指数不等于各因素指数的乘积,也就是这些因素和销售额之间无法构成指数体系,因此不能利用上述方法分析各因素的影响。

## 二、综合指数变形的平均指数体系因素分析

如果掌握的资料不足以采取综合指数体系进行因素分析,就可以根据实际情况采用综合指数变形的平均指数体系进行分析。



其原理是：

$$\begin{aligned}\frac{\sum p_1 q_1}{\sum p_0 q_0} &= \frac{\sum \left( \frac{q_1}{q_0} \cdot p_0 q_0 \right)}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum \left( \frac{q_1}{q_0} \cdot p_0 q_0 \right)} \\ &= \frac{\sum \left( \frac{p_1}{p_0} \cdot p_0 q_0 \right)}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum \left( \frac{p_1}{p_0} \cdot p_0 q_0 \right)}\end{aligned}$$

或

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{\frac{q_1}{q_0}}} \cdot \frac{\sum \frac{p_1 q_1}{\frac{q_1}{q_0}}}{\sum p_0 q_0} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{\frac{p_1}{p_0}}} \cdot \frac{\sum \frac{p_1 q_1}{\frac{p_1}{p_0}}}{\sum p_0 q_0}$$

绝对差额也存在一定的数量关系。借助上述原理，就可以进行因素分析。

某副食品商场三种商品的销售资料如表 7-6 所示。试借助指数体系，从相对数和绝对数两方面分析销售额变动及各因素的影响。

表 7-6 某副食品商场三种商品的销售资料

| 商 品 | 计 量 单 位 | 销 售 额 / 元    |               | 销售量个体指数 |
|-----|---------|--------------|---------------|---------|
|     |         | 基期 $p_0 q_0$ | 报告期 $p_1 q_1$ |         |
| 猪肉  | 千克      | 40 000       | 72 000        | 120.00% |
| 面粉  | 千克      | 31 320       | 44 640        | 106.90% |
| 洗衣粉 | 包       | 5 400        | 4 400         | 91.67%  |

指数体系的综合分析：

相对数：157.77%=112.66%×140.04%

绝对数：44 320=9 710+34 610

从以上计算结果可以看出：由于商品销售量提高 12.66% 使销售额增加 9 710 元，由于商品价格提高 40.04% 使销售额增加 34 610 元。两个因素共同作



用的结果是使商品销售额实际提高 57.77%，绝对额增加 44 320 元。

指数体系因素分析的方法也可扩展到三因素、四因素，以三因素为例，其基本原理是：

$$\frac{\sum T_1 q_1 p_1}{\sum T_0 q_0 p_0} = \frac{\sum T_1 q_0 p_0}{\sum T_0 q_0 p_0} \cdot \frac{\sum T_1 q_1 p_0}{\sum T_1 q_0 p_0} \cdot \frac{\sum T_1 q_1 p_1}{\sum T_1 q_1 p_0}$$

$$\sum T_1 q_1 p_1 - \sum T_0 q_0 p_0 = (\sum T_1 q_0 p_0 - \sum T_0 q_0 p_0) + (\sum T_1 q_1 p_0 - \sum T_1 q_0 p_0) + (\sum T_1 q_1 p_1 - \sum T_1 q_1 p_0)$$

在进行多因素分析时，要注意以下问题：

- (1) 被固定因素所属时期的选择，要遵循编制综合指数的一般原则。
- (2) 质量指标和数量指标的差别是相对的，而不能绝对化。
- (3) 对各因素的排列顺序，具体分析现象总体的经济内容，依据现象因素的联系加以具体确定。

此外，不但综合指数（包括综合指数变形的平均指数）能够构成指数体系，对平均指标的分析也可构成指数体系——可变构成指数体系。其基本原理是：

$$\frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}} = \frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_1}{\sum f_1}} \cdot \frac{\frac{\sum x_0 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}}$$

$$\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} = \frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1} + \frac{\sum x_0 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0}$$

式中， $\frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}}$  称为可变构成指数，反映了平均指标的变动程度； $\frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_1}{\sum f_1}}$  称为

固定构成指数，反映了变量值  $x$  对平均数的影响程度； $\frac{\frac{\sum x_0 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}}$  称为结构影响指

数，反映了权数  $f$  对平均数的影响程度。



通过可变构成指数体系，可以分析出平均指标的变动程度，以及两个因素（变量值和权数）对平均指标的影响程度。这里不再举例讨论。

## 任务四 掌握平均指数

### 任务卡 1

有一种说法是：“平均指数是综合指数的一种替代品。”你同意这种说法吗？为什么？

### 知识卡 1 平均指数的含义及加权算术平均指数

#### 一、平均指数的含义

平均指数是按照“先对比，后平均”的编制思路，通过对个体指数进行加权平均来计算的总指数。其实质是以个体指数作为变量，并根据个体在总体中的地位进行加权平均，以测定现象的综合变动程度。即平均指数是个体指数的加权平均数。

平均指数和综合指数都是总指数的常用编制方法，两者之间既有区别也有联系。其区别表现在：

（1）编制的思路不同，其性质表现也有所不同。综合指数是按照“先综合，后对比”的思路编制的，它直接反映现象的综合变动情况。平均指数是按照“先对比，后平均”的思路编制的，它不仅反映现象的综合变动，而且反映个体指数对总指数的影响。





(2) 使用的权数不同。综合指数以同度量因素为权数, 同度量因素可以是基期相应的数量指标或质量指标, 也可以是报告期相应的数量指标或质量指标。而平均指数以价值量指标为权数。

(3) 对资料的要求不同。编制综合指数需要全面资料, 即基期和报告期的指数化因素及同度量因素的全部资料  $p_0$ 、 $p_1$ 、 $q_0$ 、 $q_1$ , 而平均指数只需要个体指数和相应的权数资料就可以。

综合指数和平均指数之间也有联系: 两者都是总指数的编制方法, 反映的内容是一致的, 都反映了现象的综合变动程度。此外, 在特定的权数条件下, 两者之间具有变形关系。

## 二、加权算术平均指数

一般情况下, 数量指标综合指数可以改变为加权算术平均法形式计算指数, 即以数量指标个体指数为变量, 以数量指标综合指数相应的分母指标为权数, 运用加权算术平均法公式计算总指数。其计算公式是

$$\bar{K}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

因为  $k_q = \frac{q_1}{q_0}$ , 所以  $q_1 = k_q q_0$ , 有

$$\bar{K}_q = \frac{\sum k_q q_0 p_0}{\sum q_0 p_0}$$

某商场三种商品的销售量和价格如表 7-7 所示。

表 7-7 某商场三种商品的销售量和价格

| 商品 | 计量单位 | 销售量         |              | 价格 / 元   |              | 销售额 / 万元        |                  |                 |
|----|------|-------------|--------------|----------|--------------|-----------------|------------------|-----------------|
|    |      | 基期<br>$q_0$ | 报告期<br>$q_1$ | 基期 $p_0$ | 报告期<br>$p_1$ | 基期<br>$q_0 p_0$ | 报告期<br>$q_1 p_1$ | 假定<br>$q_1 p_0$ |
| 甲  | 台    | 400         | 600          | 250      | 200          | 10              | 12               | 15              |
| 乙  | 件    | 5 000       | 6 000        | 40       | 36           | 20              | 21.6             | 24              |
| 丙  | 吨    | 200         | 180          | 500      | 600          | 10              | 10.8             | 9               |
| 合计 |      |             |              |          |              | 40              | 44.4             | 48              |

现根据表 7-7 资料列出表 7-8，运用加权算术平均法指数公式计算销售量指数。

表 7-8 某商场三种商品销售量和基期销售额资料

| 商品 | 计量单位 | 销售量      |           | 基期的实际销售<br>额 $q_0p_0$ / 万元 | 个体销售量指数<br>$k_q = \frac{q_1}{q_0}$ | 假定销售额<br>$k_qq_0p_0$ / 万元 |
|----|------|----------|-----------|----------------------------|------------------------------------|---------------------------|
|    |      | 基期 $q_0$ | 报告期 $q_1$ |                            |                                    |                           |
| 甲  | 台    | 400      | 600       | 10                         | 150%                               | 15                        |
| 乙  | 件    | 5 000    | 6 000     | 20                         | 120%                               | 24                        |
| 丙  | 吨    | 200      | 180       | 10                         | 90%                                | 9                         |
| 合计 |      |          |           | 40                         |                                    | 48                        |

销售量指数为

$$\bar{K}_q = \frac{\sum k_q q_0 p_0}{\sum q_0 p_0} = \frac{48}{40} \times 100\% = 120\%$$

$$\sum k_q q_0 p_0 - \sum q_0 p_0 = 48 - 40 = 8 \text{ (万元)}$$

显然，采用加权算术平均法总指数公式计算的销售量指数与采用综合法总指数公式计算的销售量指数结果是一致的，它们的经济意义也完全相同。

任务卡 2

你了解加权调和平均指数吗？请查找资料，写下你的收获。

知识卡 2 加权调和平均指数

一般情况下，质量指标综合指数可以改变为加权调和平均法形式计算指数，即以质量指标个体指数的倒数为变量，以质量指标综合指数相应的分子指标为权数，运用加权调和平均法公式计算总指数。其计算公式为



$$\bar{K}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

因为  $k_p = \frac{p_1}{p_0}$ , 所以  $\frac{1}{k_p} = \frac{p_0}{p_1}$ , 有

$$\bar{K}_p = \frac{\sum p_1 q_1}{\sum \frac{1}{k_p} p_1 q_1}$$

现根据表 7-7 资料列出表 7-9, 运用加权调和平均法总指数公式计算销售价格指数。

表 7-9 某商场三种商品的销售量和报告期销售额

| 商品 | 计量单位 | 价格 / 元   |           | 报告期的实际销售额 $p_1 q_1$ / 万元 | 个体价格指数 $k_p = \frac{p_1}{p_0}$ | 假定销售额 $\frac{1}{k_p} p_1 q_1$ / 万元 |
|----|------|----------|-----------|--------------------------|--------------------------------|------------------------------------|
|    |      | 基期 $p_0$ | 报告期 $p_1$ |                          |                                |                                    |
| 甲  | 台    | 250      | 200       | 12                       | 80%                            | 15                                 |
| 乙  | 件    | 40       | 36        | 21.6                     | 90%                            | 24                                 |
| 丙  | 吨    | 500      | 600       | 10.8                     | 120%                           | 9                                  |
| 合计 |      |          |           | 44.4                     |                                | 48                                 |

显然, 采用加权调和平均法总指数公式计算的销售价格指数与采用综合法总指数公式计算的销售价格指数结果同样一致, 它们的经济意义也完全相同。

$$\bar{K}_p = \frac{\sum p_1 q_1}{\sum \frac{1}{k_p} p_1 q_1} = \frac{44.4}{48} \times 100\% = 92.5\%$$

$$\sum p_1 q_1 - \sum \frac{1}{k_p} p_1 q_1 = 44.4 - 48 = -3.6 \text{ (万元)}$$



## 职业基础知识测试

### 一、填空题

1. 广义的指数是用以测定某个变量在时间或空间上\_\_\_\_\_的相对数。它除了包括狭义的指数外, 还包括计划完成程度相对数、动态相对数等。狭义的指



数是一种特殊的相对数，它是说明由许多不能直接加总的要素所组成的\_\_\_\_\_综合变动的相对数。

2. 统计指数的作用包括综合反映复杂现象的总体数量变动的\_\_\_\_\_、程度和\_\_\_\_\_, 对现象\_\_\_\_\_进行因素分析。

3. 广义的指数体系类似于指标体系的概念，泛指由若干个\_\_\_\_\_的统计指数所结成的体系。狭义的指数体系仅指几个指数在一定的经济基础上所结成的\_\_\_\_\_的数量关系式。

4. 平均指数是以个体指数为变量，并根据个体在总体中的地位进行\_\_\_\_\_, 来测定现象的\_\_\_\_\_, 即平均指数是个体指数的\_\_\_\_\_。

## 二、简答题

1. 简述统计指数的分类方法和种类。
2. 什么是综合指数？
3. 简述综合指数的编制原理。
4. 比较数量指标和质量指标的异同。
5. 指数体系有哪些作用？



## 职业核心能力训练

### 【实训目标】

(1) 深刻理解并掌握居民消费价格指数、商品零售价格指数、居民货币购买力指数、通货膨胀（通货紧缩）指数的概念及彼此之间的关系。

(2) 及时了解全国、本地国民经济的发展现状。

### 【具体要求】

通过互联网、专业期刊、图书查询我国及你所在的省（直辖市、自治区、特别行政区）近五年的居民消费价格指数、商品零售价格指数，计算出居民货币购买力指数和通货膨胀（通货紧缩）指数，并写出分析报告。

### 【成绩评定】

根据分析报告的全面性、科学性、深入性综合评定。



## 【实训记录】

## 【交流与评价】

| 项 目  | 成 绩                 |                     |                     |     |
|------|---------------------|---------------------|---------------------|-----|
|      | 收集资料的真实性<br>( 30% ) | 收集资料的全面性<br>( 30% ) | 分析报告的合理性<br>( 40% ) | 总 分 |
| 自我评价 |                     |                     |                     |     |
| 同学评价 |                     |                     |                     |     |
| 教师评价 |                     |                     |                     |     |

## 附录 正态分布概率表

| $t$  | $F(t)$ | $t$  | $F(t)$ | $t$  | $F(t)$ | $t$  | $F(t)$ |
|------|--------|------|--------|------|--------|------|--------|
| 0.00 | 0.0000 | 0.32 | 0.2510 | 0.64 | 0.4778 | 0.96 | 0.6929 |
| 0.01 | 0.0080 | 0.33 | 0.2586 | 0.65 | 0.4843 | 0.97 | 0.6680 |
| 0.02 | 0.0160 | 0.34 | 0.2661 | 0.66 | 0.4907 | 0.98 | 0.6729 |
| 0.03 | 0.0239 | 0.35 | 0.2737 | 0.67 | 0.4971 | 0.99 | 0.6778 |
| 0.04 | 0.0319 | 0.36 | 0.2812 | 0.68 | 0.5035 | 1.00 | 0.6827 |
| 0.05 | 0.0399 | 0.37 | 0.2886 | 0.69 | 0.5098 | 1.01 | 0.6875 |
| 0.06 | 0.0478 | 0.38 | 0.2961 | 0.70 | 0.5161 | 1.02 | 0.6923 |
| 0.07 | 0.0558 | 0.39 | 0.3035 | 0.71 | 0.5223 | 1.03 | 0.6970 |
| 0.08 | 0.0638 | 0.40 | 0.3108 | 0.72 | 0.5285 | 1.04 | 0.7017 |
| 0.09 | 0.0717 | 0.41 | 0.3182 | 0.73 | 0.5346 | 1.05 | 0.7063 |
| 0.10 | 0.0797 | 0.42 | 0.3255 | 0.74 | 0.5407 | 1.06 | 0.7109 |
| 0.11 | 0.0876 | 0.43 | 0.3328 | 0.75 | 0.5467 | 1.07 | 0.7154 |
| 0.12 | 0.0955 | 0.44 | 0.3401 | 0.76 | 0.5527 | 1.08 | 0.7199 |
| 0.13 | 0.0103 | 0.45 | 0.3473 | 0.77 | 0.5587 | 1.09 | 0.7243 |
| 0.14 | 0.0111 | 0.46 | 0.3545 | 0.78 | 0.5646 | 1.10 | 0.7287 |
| 0.15 | 0.1192 | 0.47 | 0.3616 | 0.79 | 0.5705 | 1.11 | 0.7330 |
| 0.16 | 0.1271 | 0.48 | 0.3688 | 0.80 | 0.5763 | 1.12 | 0.7373 |
| 0.17 | 0.1350 | 0.49 | 0.3759 | 0.81 | 0.5821 | 1.13 | 0.7415 |
| 0.18 | 0.1428 | 0.50 | 0.3829 | 0.82 | 0.5878 | 1.14 | 0.7457 |
| 0.19 | 0.1507 | 0.51 | 0.3899 | 0.83 | 0.5935 | 1.15 | 0.7499 |
| 0.20 | 0.1585 | 0.52 | 0.3969 | 0.84 | 0.5991 | 1.16 | 0.7540 |
| 0.21 | 0.1663 | 0.53 | 0.4039 | 0.85 | 0.6047 | 1.17 | 0.7580 |
| 0.22 | 0.1741 | 0.54 | 0.4108 | 0.86 | 0.6102 | 1.18 | 0.7620 |
| 0.23 | 0.1819 | 0.55 | 0.4177 | 0.87 | 0.6157 | 1.19 | 0.7660 |
| 0.24 | 0.1897 | 0.56 | 0.4245 | 0.88 | 0.6211 | 1.20 | 0.7699 |
| 0.25 | 0.1974 | 0.57 | 0.4313 | 0.89 | 0.6265 | 1.21 | 0.7737 |
| 0.26 | 0.2051 | 0.58 | 0.4381 | 0.90 | 0.6319 | 1.22 | 0.7775 |
| 0.27 | 0.2128 | 0.59 | 0.4448 | 0.91 | 0.6372 | 1.23 | 0.7813 |
| 0.28 | 0.2205 | 0.60 | 0.4515 | 0.92 | 0.6424 | 1.24 | 0.7850 |
| 0.29 | 0.2282 | 0.61 | 0.4581 | 0.93 | 0.6476 | 1.25 | 0.7887 |
| 0.30 | 0.2358 | 0.62 | 0.4647 | 0.94 | 0.6528 | 1.26 | 0.7923 |
| 0.31 | 0.2434 | 0.63 | 0.4713 | 0.95 | 0.6579 | 1.27 | 0.7959 |



(续表)

| $t$  | $F(t)$ | $t$  | $F(t)$ | $t$  | $F(t)$ | $t$  | $F(t)$   |
|------|--------|------|--------|------|--------|------|----------|
| 1.28 | 0.7995 | 1.61 | 0.8926 | 1.94 | 0.9476 | 2.54 | 0.9889   |
| 1.29 | 0.8030 | 1.62 | 0.8948 | 1.95 | 0.9488 | 2.56 | 0.9895   |
| 1.30 | 0.8064 | 1.63 | 0.8969 | 1.96 | 0.9500 | 2.58 | 0.9901   |
| 1.31 | 0.8098 | 1.64 | 0.8990 | 1.97 | 0.9512 | 2.60 | 0.9907   |
| 1.32 | 0.8132 | 1.65 | 0.9011 | 1.98 | 0.9523 | 2.62 | 0.9912   |
| 1.33 | 0.8165 | 1.66 | 0.9031 | 1.99 | 0.9534 | 2.64 | 0.9917   |
| 1.34 | 0.8198 | 1.67 | 0.9051 | 2.00 | 0.9545 | 2.66 | 0.9922   |
| 1.35 | 0.8230 | 1.68 | 0.9070 | 2.02 | 0.9566 | 2.68 | 0.9926   |
| 1.36 | 0.8262 | 1.69 | 0.9090 | 2.04 | 0.9587 | 2.70 | 0.9931   |
| 1.37 | 0.8293 | 1.70 | 0.9109 | 2.06 | 0.9606 | 2.72 | 0.9935   |
| 1.38 | 0.8324 | 1.71 | 0.9127 | 2.08 | 0.9625 | 2.74 | 0.9939   |
| 1.39 | 0.8355 | 1.72 | 0.9146 | 2.10 | 0.9643 | 2.76 | 0.9942   |
| 1.40 | 0.8385 | 1.73 | 0.9164 | 2.12 | 0.9660 | 2.78 | 0.9946   |
| 1.41 | 0.8415 | 1.74 | 0.9181 | 2.14 | 0.9676 | 2.80 | 0.9949   |
| 1.42 | 0.8444 | 1.75 | 0.9199 | 2.16 | 0.9692 | 2.82 | 0.9952   |
| 1.43 | 0.8473 | 1.76 | 0.9216 | 2.18 | 0.9707 | 2.84 | 0.9955   |
| 1.44 | 0.8501 | 1.77 | 0.9233 | 2.20 | 0.9722 | 2.86 | 0.9958   |
| 1.45 | 0.8529 | 1.78 | 0.9249 | 2.22 | 0.9736 | 2.88 | 0.9960   |
| 1.46 | 0.8557 | 1.79 | 0.9265 | 2.24 | 0.9749 | 2.90 | 0.9962   |
| 1.47 | 0.8584 | 1.80 | 0.9281 | 2.26 | 0.9762 | 2.92 | 0.9965   |
| 1.48 | 0.8611 | 1.81 | 0.9297 | 2.28 | 0.9774 | 2.94 | 0.9967   |
| 1.49 | 0.8638 | 1.82 | 0.9312 | 2.30 | 0.9786 | 2.96 | 0.9969   |
| 1.50 | 0.8664 | 1.83 | 0.9328 | 2.32 | 0.9797 | 2.98 | 0.9971   |
| 1.51 | 0.8690 | 1.84 | 0.9342 | 2.34 | 0.9807 | 3.00 | 0.9973   |
| 1.52 | 0.8715 | 1.85 | 0.9357 | 2.36 | 0.9817 | 3.20 | 0.9986   |
| 1.53 | 0.8740 | 1.86 | 0.9371 | 2.38 | 0.9827 | 3.40 | 0.9993   |
| 1.54 | 0.8764 | 1.87 | 0.9385 | 2.40 | 0.9836 | 3.60 | 0.99968  |
| 1.55 | 0.8789 | 1.88 | 0.9399 | 2.42 | 0.9845 | 3.80 | 0.99986  |
| 1.56 | 0.8812 | 1.89 | 0.9412 | 2.44 | 0.9853 | 4.00 | 0.99994  |
| 1.57 | 0.8836 | 1.90 | 0.9426 | 2.46 | 0.9861 | 4.50 | 0.999993 |
| 1.58 | 0.8859 | 1.91 | 0.9439 | 2.48 | 0.9869 | 5.00 | 0.999999 |
| 1.59 | 0.8882 | 1.92 | 0.9451 | 2.50 | 0.9876 |      |          |
| 1.60 | 0.8904 | 1.93 | 0.9464 | 2.52 | 0.9883 |      |          |



## 参考文献

- [1] 吴喜之, 吕晓玲. 统计学: 从数据到结论 [M]. 5 版. 北京: 中国统计出版社, 2021.
- [2] 贾俊平, 何晓群, 金勇进. 统计学 [M]. 8 版. 北京: 中国人民大学出版社, 2021.
- [3] 贾俊平. 统计学基础 [M]. 6 版. 北京: 中国人民大学出版社, 2021.
- [4] 洪永淼. 概率论与统计学 [M]. 2 版. 北京: 中国统计出版社, 2021.
- [5] 盛骤, 谢式千, 潘承毅. 概率论与数理统计 [M]. 5 版. 北京: 高等教育出版社, 2019.
- [6] 李航. 统计学习方法 [M]. 2 版. 北京: 清华大学出版社, 2019.
- [7] 刘强. 大数据时代的统计学思维: 让你从众多数据中找到真相 [M]. 北京: 中国水利水电出版社, 2018.