

21世纪高等教育标准教材

STATISTICS

# 统计学

( 第二版 )

冯力 主编

21世纪高等教育标准教材

STATISTICS

# 统计学

第二版 )

冯力 主编

 东北财经大学出版社  
Dongbei University of Finance & Economics Press

大连

© 冯力 2015

图书在版编目 (CIP) 数据

统计学 / 冯力主编. —2 版. —大连: 东北财经大学出版社, 2015. 2

(21 世纪高等教育标准教材)

ISBN 978-7-5654-1814-3

I. 统… II. 冯… III. 统计学-高等学校-教材 IV. C8

中国版本图书馆 CIP 数据核字 (2015) 第 016120 号

东北财经大学出版社出版

(大连市黑石礁尖山街 217 号 邮政编码 116025)

教学支持: (0411) 84710309

营 销 部: (0411) 84710711

总 编 室: (0411) 84710523

网 址: <http://www.dufep.cn>

读者信箱: [dufep@dufe.edu.cn](mailto:dufep@dufe.edu.cn)

大连图腾彩色印刷有限公司印刷

东北财经大学出版社发行

幅面尺寸: 170mm×240mm

字数: 405 千字

印张: 19 1/2

2015 年 2 月第 2 版

2015 年 2 月第 6 次印刷

责任编辑: 时 博 王芃南

责任校对: 王 娟 那 欣 赵 楠

封面设计: 潘 凯

版式设计: 钟福建

定价: 32.00 元

## 第二版前言

《统计学》自2010年第一版出版以来，得到了广大读者的认可，销量不断增加，这对我们是莫大的鼓励，也增加了我们不断完善这本书的信心。第二版在内容体系上与第一版相似，依旧本着方法原理、软件操作与案例分析紧密结合，方法相对独立，小步循环的原则进行编排，但对某些章节的机上作业进行了更换和调整。此外，也订正了一些过时的信息，修正了一些不规范的表述，并对全书的公式进行了仔细核查，改正了第一版中存在的个别错误。

第二版教材编写分工情况如下：第1章、第4章、第7章、第8章、第9章、第10章、第11章由冯力教授编写；第2章由刘沈忠副教授编写；第3章由冯叔民副教授编写；第5章由孙玉环副教授编写；第6章由马晓君讲师编写。全书由冯力教授主编，徐建邦教授主审，感谢东北财经大学出版社对本书出版的大力支持。

虽然已经竭尽全力，但由于水平有限，不免有错误和疏漏之处，敬请读者提出宝贵意见，不胜感激。

编者

2014年12月

# 第一版前言

本教材是在参考国内外优秀统计教材的基础上，结合财经类院校的教学特点编写的，主要面向具有一定概率论与数理统计基础的非数学专业的在校本科生以及实际从事数据分析活动的工作人员。

我们将整个统计方法体系的构成内容区分为基础统计与高级统计两部分，针对单变量或双变量截面数据的描述与推断方法属于基础统计。本教材集中讲解基础统计的方法原理。

如果离开了统计软件，统计方法在实际工作中几乎没有用武之地，将统计方法原理讲解与统计软件实际操作结合起来展开统计教学，一直是高等学校统计教学改革的重要目标。本教材引入了 SPSS 的有关操作方法作为教学内容的有机组成部分，力求避免以往原理教学与软件教学相互脱节的不利状况。此外，本教材中还增添了实验教学的内容，以体现理论教学与实践教学相结合的宗旨。

教材中各个章节内容的编排遵循这样一条循序渐进的原则：引例→方法原理→软件操作→问题思考→机上作业→实验课题。与此同时，特别强调方法原理的准确理解与数据处理的熟练操作。

全书共分 11 章。第 1 章、第 4 章、第 8 章、第 9 章、第 10 章、第 11 章由冯力编写；第 2 章由刘沈忠编写；第 3 章由冯叔民编写；第 5 章由孙玉环编写；第 6 章由马晓君编写；第 7 章由庄连平编写。全书由冯力主编，徐建邦主审，冯力对全书进行了总纂、定稿。

由于时间和水平有限，错误与疏漏在所难免，敬请读者提出宝贵意见，以利于统计教材和统计教学的不断改进。

编者

2010 年 11 月

# 目 录

|                       |    |
|-----------------------|----|
| 第 1 章 总论              | 1  |
| 引例 1                  | 1  |
| 1.1 统计中的常用术语          | 2  |
| 1.2 统计活动的基本程序         | 5  |
| 1.3 统计方法的基本构成         | 6  |
| 1.4 样本数据的基础结构         | 10 |
| 1.5 统计中的误差            | 12 |
| 1.6 将样本数据转换为 SPSS 数据集 | 13 |
| 本章小结                  | 23 |
| 问题思考                  | 23 |
| 机上作业                  | 24 |
| 实验课题                  | 27 |
| 附录 1.1 SPSS 数据预处理     | 29 |
| 附录 1.2 问卷调查法          | 36 |
| 第 2 章 图表描述            | 41 |
| 引例 2                  | 41 |
| 2.1 品质型数据的图表描述        | 42 |
| 2.2 数值型数据的图表描述        | 46 |
| 2.3 双变量关系的图表描述        | 53 |
| 2.4 运用 SPSS 进行图表描述    | 55 |
| 本章小结                  | 67 |
| 问题思考                  | 68 |
| 机上作业                  | 68 |
| 实验课题                  | 70 |
| 第 3 章 描述性统计量          | 72 |
| 引例 3                  | 72 |
| 3.1 集中趋势描述性统计量        | 73 |
| 3.2 离散程度描述性统计量        | 78 |
| 3.3 分布形态描述性统计量        | 84 |
| 3.4 运用 SPSS 进行统计量描述   | 85 |
| 本章小结                  | 91 |
| 问题思考                  | 91 |

|                           |     |
|---------------------------|-----|
| 机上作业 .....                | 91  |
| 实验课题 .....                | 94  |
| <b>第 4 章 参数估计</b> .....   | 96  |
| 引例 4 .....                | 96  |
| 4.1 抽样分布 .....            | 97  |
| 4.2 点估计 .....             | 100 |
| 4.3 单样本总体均值区间估计 .....     | 102 |
| 4.4 两个独立样本总体均值差区间估计 ..... | 105 |
| 4.5 两个匹配样本总体均值差区间估计 ..... | 108 |
| 4.6 样本容量的确定 .....         | 110 |
| 4.7 总体方差区间估计 .....        | 112 |
| 4.8 总体比率区间估计 .....        | 114 |
| 4.9 运用 SPSS 进行参数估计 .....  | 117 |
| 本章小结 .....                | 120 |
| 问题思考 .....                | 121 |
| 机上作业 .....                | 122 |
| 实验课题 .....                | 124 |
| 附录 4.1 几种重要的概率分布 .....    | 125 |
| <b>第 5 章 假设检验</b> .....   | 130 |
| 引例 5 .....                | 130 |
| 5.1 假设检验基本原理 .....        | 130 |
| 5.2 单样本均值检验 .....         | 138 |
| 5.3 两个独立样本均值差检验 .....     | 141 |
| 5.4 两个匹配样本均值差检验 .....     | 144 |
| 5.5 总体方差假设检验 .....        | 145 |
| 5.6 总体比率假设检验 .....        | 147 |
| 5.7 运用 SPSS 进行假设检验 .....  | 149 |
| 本章小结 .....                | 153 |
| 问题思考 .....                | 154 |
| 机上作业 .....                | 154 |
| 实验课题 .....                | 156 |
| <b>第 6 章 方差分析</b> .....   | 158 |
| 引例 6 .....                | 158 |
| 6.1 方差分析的理论假设 .....       | 159 |
| 6.2 方差分析的基本思想和基本步骤 .....  | 160 |
| 6.3 方差相等性检验 .....         | 165 |
| 6.4 方差分析中的多重比较 .....      | 166 |

|                             |     |
|-----------------------------|-----|
| 6.5 试验数据与观察数据 .....         | 169 |
| 6.6 运用 SPSS 进行方差分析 .....    | 171 |
| 本章小结 .....                  | 175 |
| 问题思考 .....                  | 176 |
| 机上作业 .....                  | 176 |
| 实验课题 .....                  | 179 |
| <b>第 7 章 列联分析</b> .....     | 180 |
| 引例 7 .....                  | 180 |
| 7.1 列联表 .....               | 181 |
| 7.2 列联表中的 $\chi^2$ 检验 ..... | 183 |
| 7.3 列联表中的相关系数 .....         | 186 |
| 7.4 运用 SPSS 进行列联分析 .....    | 188 |
| 本章小结 .....                  | 192 |
| 问题思考 .....                  | 193 |
| 机上作业 .....                  | 193 |
| 实验课题 .....                  | 196 |
| <b>第 8 章 回归分析</b> .....     | 198 |
| 引例 8 .....                  | 198 |
| 8.1 相关系数 .....              | 199 |
| 8.2 回归模型与回归方程 .....         | 204 |
| 8.3 估计的回归方程 .....           | 206 |
| 8.4 判定系数 .....              | 208 |
| 8.5 $F$ 检验 .....            | 211 |
| 8.6 回归预测 .....              | 212 |
| 8.7 残差分析 .....              | 218 |
| 8.8 运用 SPSS 进行回归分析 .....    | 224 |
| 本章小结 .....                  | 229 |
| 问题思考 .....                  | 230 |
| 机上作业 .....                  | 231 |
| 实验课题 .....                  | 234 |
| 附录 8.1 曲线估计 .....           | 236 |
| <b>第 9 章 定序数据分析</b> .....   | 239 |
| 引例 9 .....                  | 239 |
| 9.1 威尔科克森秩和检验 .....         | 240 |
| 9.2 符号检验 .....              | 245 |
| 9.3 威尔科克森符号秩和检验 .....       | 248 |
| 9.4 运用 SPSS 进行定序数据分析 .....  | 252 |

|                         |     |
|-------------------------|-----|
| 本章小结 .....              | 255 |
| 问题思考 .....              | 256 |
| 机上作业 .....              | 257 |
| 实验课题 .....              | 258 |
| 第 10 章 指数 .....         | 260 |
| 引例 10 .....             | 260 |
| 10.1 综合指数 .....         | 261 |
| 10.2 平均指数 .....         | 264 |
| 10.3 关于总指数计算方法的修正 ..... | 265 |
| 10.4 指数体系 .....         | 268 |
| 10.5 经济统计中几种常见的指数 ..... | 269 |
| 本章小结 .....              | 271 |
| 问题思考 .....              | 272 |
| 机上作业 .....              | 272 |
| 第 11 章 时间序列 .....       | 275 |
| 引例 11 .....             | 275 |
| 11.1 时间序列构成 .....       | 276 |
| 11.2 波动性的描述 .....       | 278 |
| 11.3 长期趋势的测定 .....      | 280 |
| 11.4 季节波动的测定 .....      | 284 |
| 本章小结 .....              | 288 |
| 问题思考 .....              | 289 |
| 机上作业 .....              | 290 |
| 主要参考文献 .....            | 293 |
| 附录 常用统计表 .....          | 294 |

# 第1章 总论

## 引例 1

一家灯泡生产企业采用新技术生产一批具备节能、亮度大、外形美观等诸多优点的新型灯泡。为了顺利打开销售市场，除上述优势之外，管理者觉得还有必要就这种新型灯泡的使用寿命，向消费者作出明确承诺。但灯泡使用寿命的测试是具有破坏性的，显然无法针对每一只灯泡直接给出测试结果。这正是统计方法的用武之地。

统计人员从全部灯泡产品中随机抽取了 100 只，经过测试获得如下数据（见表 1-1）：

表 1-1 100 只灯泡使用寿命测试结果（小时）

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 693 | 717 | 665 | 719 | 685 | 709 | 691 | 688 | 705 | 718 |
| 706 | 715 | 712 | 722 | 691 | 708 | 690 | 692 | 707 | 701 |
| 664 | 708 | 729 | 694 | 668 | 674 | 692 | 710 | 696 | 702 |
| 681 | 681 | 661 | 695 | 710 | 658 | 691 | 722 | 651 | 741 |
| 721 | 685 | 735 | 691 | 693 | 695 | 747 | 694 | 673 | 698 |
| 720 | 716 | 728 | 713 | 697 | 666 | 699 | 690 | 749 | 713 |
| 677 | 706 | 733 | 699 | 726 | 684 | 700 | 736 | 708 | 676 |
| 679 | 718 | 712 | 725 | 704 | 696 | 682 | 689 | 727 | 702 |
| 695 | 707 | 683 | 696 | 729 | 698 | 698 | 683 | 688 | 701 |
| 697 | 683 | 692 | 717 | 703 | 706 | 700 | 685 | 689 | 671 |

为便于整体观察，这些数据又被整理成表 1-2 和图 1-1 的形式。

表 1-2 100 只灯泡使用寿命频数分布表

| 使用寿命（小时）  | 频数（只） |
|-----------|-------|
| 650 ~ 661 | 2     |
| 661 ~ 672 | 6     |
| 672 ~ 683 | 8     |
| 683 ~ 694 | 22    |
| 694 ~ 705 | 23    |
| 705 ~ 716 | 17    |
| 716 ~ 727 | 12    |
| 727 ~ 738 | 7     |
| 738 ~ 749 | 2     |
| 749 ~ 751 | 1     |
| 合 计       | 100   |

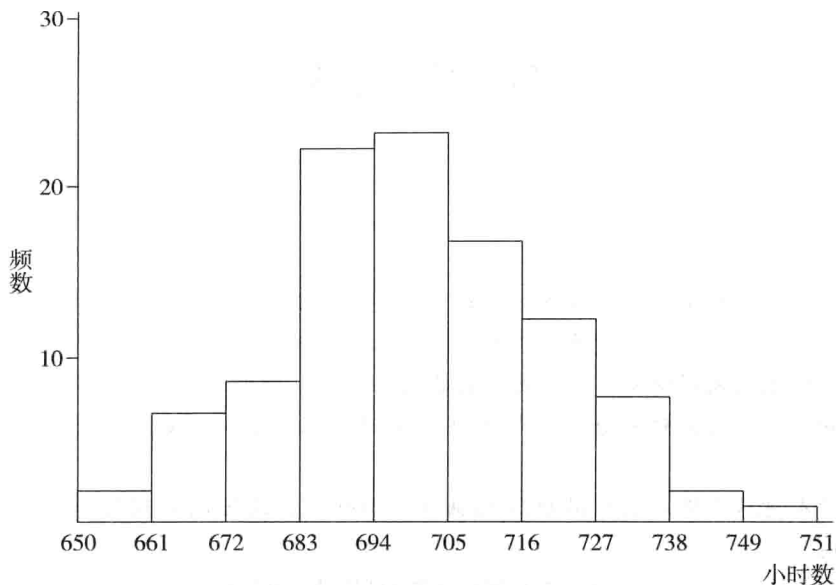


图 1-1 100 只灯泡使用寿命频数分布图

从表 1-2、图 1-1 中可以看出, 100 只灯泡的使用寿命最低不低于 650 小时; 最高不高于 751 小时; 大多数灯泡的使用寿命在 683 小时 ~ 727 小时之间。经过计算得知, 100 只灯泡的平均使用寿命为 699.85 小时。

统计人员依据统计学的方法原理进一步作出估计: 全部灯泡产品的平均使用寿命应该在 695.95 小时 ~ 703.75 小时之间, 并且承诺, 作出这种估计的把握程度为 95%。

如您所见, 统计人员在此过程中并没能给出全部灯泡产品中, 每一只灯泡使用寿命的确切数值, 这或许是一种遗憾, 但他们在有一定把握的情况下, 给出了全部灯泡产品使用寿命的一般水平及其存在区间, 从而为企业产品销售策略的制定提供了比较可靠的数字依据。

除此之外, 您还能想出更好的方法来了解和确认这批灯泡产品的使用寿命吗? 在没有找到更好的方法之前, 不妨先来学习统计方法。

## 1.1 统计中的常用术语

统计是专门研究如何搜集数据、整理数据、分析数据, 以获取研究对象总体有用信息的一门方法科学。为便于今后有关统计方法、原理的系统讲解, 有必要先来熟悉一下统计中的几个常用术语。

### 1.1.1 总体

统计方法的有效运用建立在研究对象的大量性的基础之上。假如我们的研究对象仅仅是一两只灯泡, 那就没有必要采用统计方法来确定其使用寿命。统计中习惯

将研究对象中所包含的大量个体的全体称为总体，将总体中所包含的个体的数量称为总体容量。

根据我们感兴趣的对象不同，总体可以是刚刚生产出来的一批灯泡，也可以是一群人，还可以是众多的企业等。总体容量可以是有限的，也可以是无限的。

一个总体的确定依赖于我们最初的研究目的，研究目的一经确定，相应的总体也就唯一地确定了。我们的研究目的是要确认全部灯泡产品的使用寿命，那么这批灯泡就构成了一个总体。研究目的变了，是要了解全体在校学生的身体健康情况，那么这一群在校学生就构成了另一个总体。

通常情况下，我们所面对的多为有限总体。但在统计学家那里，他们更喜欢无限总体，因为，对于无限总体，现成的数学方法可以直接加以引用，而不必为分析结果只是一种近似而心有不安。但如果某个有限总体所包含的个体数目足够大，此时将它看做无限总体用数学方法加以处理也应该算做是一种有效的“近似”，结果仍然是可取的。

同一个总体概念，在数学家看来情形又有所不同，在数学家眼里没有所谓个体，他们把某一变量的所有可能取值看做是一个总体，这与我们给出的总体概念是不矛盾的。

### 1.1.2 总体参数

统计研究的直接目的就是要确认总体在某一个方面的数量特征，譬如全部灯泡产品的平均使用寿命、合格率、使用寿命长短的差异程度等；再譬如全体在校学生的平均身高、平均体重、身高或体重的差异程度等。总体的这些需要我们加以确认的某个方面的数量特征在统计中称做总体参数。

统计过程的最终成果就是获取了总体参数的具体取值。由于总体是唯一确定下来的，所以总体参数的具体取值也将是唯一确定的，从道理上讲，它可以通过总体数据来计算得出，但总体数据往往是无法掌握的，也正是由于这个原因，才需要我们运用统计方法来寻求总体参数的这个未知但却是唯一确定的数字。

### 1.1.3 样本

统计方法的运用带有迂回性的特点，既然总体数据无法或不便于直接获取，那么就可以先从总体中抽取一部分个体，就这些个体来搜集有关数据，进而间接地估计出总体参数的具体取值。统计中将达到对总体参数认识的目的，而从总体中抽取出来的部分个体称做样本，将样本中所包含的个体的数量称做样本容量。总体容量可以是有限的，也可以是无限的，但样本容量却总是有限的，而且通常是远远小于总体容量的。

为了保证对总体参数的推断结果符合给定的把握程度和准确程度，样本的抽取一般要遵循随机原则，即样本中的每一个体都是随机地获得的。与总体的唯一性和确定性形成鲜明对照，样本却是带有随机性的，我们最终获取的样本只是在随机原则背景下存在着的众多或无数个可能样本中的一个。

#### 1.1.4 样本统计量

简言之,由样本数据加工计算出来的用以推断总体参数的数,称做统计量。样本及样本数据的获取是带有随机性的,因此,由样本数据计算出来的统计量的值同样也带有随机性,或者说,样本统计量是随机变量。所谓统计推断,实际上就是用带有随机性的样本统计量的值来推断未知的、然而却是唯一确定的总体参数的值。

#### 1.1.5 变量及其观测尺度

与最初的研究目的相关联,我们总是对总体及其所包含的个体身上的某一个或某几个方面的属性感兴趣,这些属性在统计中又被称做变量。总体中每一个个体的属性是可以无限列举的,构成统计中变量的只是那些我们感兴趣的属性。譬如灯泡的“使用寿命”、“质量等级”;在校学生的“性别”、“身高”、“体重”等,都可以成为统计研究中的变量。

变量可以根据获取该变量观测值时所采用的尺度不同来进行类型划分。变量的观测尺度可以区分为四种类型,即定类尺度、定序尺度、定距尺度和定比尺度,而且仅仅是这四种尺度,没有第五种尺度。

##### 1) 定类尺度

总体或样本中每一个个体身上与某一变量相对应的那种属性的具体特征,有时可以直接用数字来刻画和衡量,有时则无法直接用数字来刻画和衡量。定类尺度正适用于这种其属性特征无法用数字直接刻画和衡量情况下的观测值的测量。这是一种最为粗糙的测量尺度,只能对总体或样本中所包含的个体进行平行分类,而且各类别之间不存在顺序上的差异。

譬如,性别这个变量,其观测值就只能采用定类尺度来测量,其具体的观测值为“男”或“女”。此外,像“民族”、“行政区划”、“职业”等,也都是适合采用定类尺度来获取观测值的变量。

由定类尺度所获得的数据,只能进行是非判断的运算,而不能直接进行大小比较、加减、乘除的运算。

##### 2) 定序尺度

与定类尺度相比较,定序尺度的测量结果就要来得较为精确一些,其观测值不仅可以反映个体之间类别上的不同,同时也能够反映不同类别之间顺序上的差异。

譬如,“产品等级”这个变量,采用定序尺度所获取的观测值为“一等品”、“二等品”和“三等品”,不同等级的产品分属不同的类别,同时各类别之间又存在着顺序上的差异,从“一等品”到“二等品”再到“三等品”,产品质量越来越低。再譬如,“学习成绩”这个变量,其观测值从“优”到“良”再到“及格”最后到“不及格”,不仅反映学生的学习成绩不尽相同,同时也反映成绩越来越差。

由定序尺度所获得的数据,不仅可以进行是非判断的运算,而且也可以进行大小比较的运算,但不能进行加减、乘除运算。

### 3) 定距尺度

与定序尺度相比较,定距尺度观测结果的精确程度又进了一步,其观测值可以直接表现为数字。这种表现为数字的观测值不仅可以反映个体之间类别的不同、顺序的差异,而且还可以反映差异的多少。

“时间”和“温度”是两个典型的适合采用定距尺度获取观测值的变量。由定距尺度所获得的数据,不仅可以进行是非判断、大小比较的运算,而且还可以进行加减运算,但不能进行乘除运算。从“8点钟”到“16点钟”,两个不同的时间相差8小时,但如果说“16点钟”是“8点钟”的2倍,就没有意义了。

### 4) 定比尺度

与前3种观测尺度相比较,定比尺度是最为精确的观测尺度,其观测值也直接表现为数字。实践中我们所遇到的绝大多数变量都属于适合采用这类尺度获取观测值的变量,譬如,“灯泡使用寿命”、“年龄”、“身高”、“体重”等。由定比尺度所获取的数据不仅适合是非判断、大小比较和加减运算,还可以进行乘除运算。

依据观测尺度不同,可以将变量划分为定类变量、定序变量、定距变量和定比变量四种类型。实践中有时也将定类变量与定序变量合为一类称做品质型变量;将定距变量与定比变量合为一类称做数值型变量。一般的统计软件中则是将变量区分为定类变量、定序变量和数值型变量3种类型。表1-3概括了变量类型划分的具体情况。

表 1-3

变量类型的划分

| 变量<br>尺度 | 品质型变量 |      | 数值型变量 |      |
|----------|-------|------|-------|------|
|          | 定类变量  | 定序变量 | 定距变量  | 定比变量 |
| 定类 (=、≠) | √     | √    | √     | √    |
| 定序 (<、>) |       | √    | √     | √    |
| 定距 (+、-) |       |      | √     | √    |
| 定比 (×、÷) |       |      |       | √    |

## 1.2 统计活动的基本程序

尽管在实际的统计活动中,由于研究对象不同,具体的统计内容会有千差万别,但作为从数量方面认识事物的一种专门方法,其整个操作过程还是有着相对固定的步骤和程序的。统计活动的基本程序可用图1-2来概括。

统计活动的全部过程是围绕着总体逐步展开的,总体参数是统计认识的最终目标,但实际的操作过程却是绕了一个圈子,走的是一条迂回路线。由总体按照特定方法抽取样本,就变量针对样本中的每一个个体进行观察和测量得到样本数据,由样本数据计算出统计量的值,再由统计量的值来估计总体参数,从而完成了一个完整的统计推断过程。

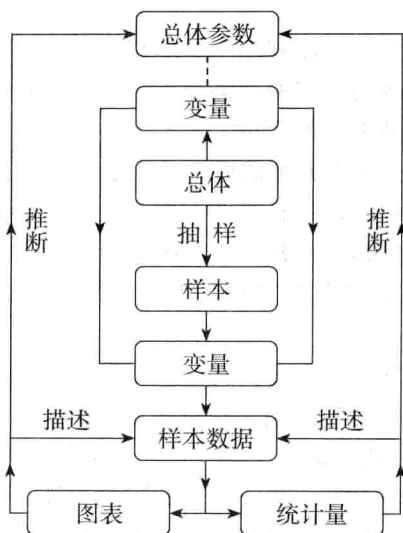


图 1-2 统计活动的基本程序

原始的样本数据通常是大量、零散、杂乱无章、令人眼花缭乱的，根本无法直接对其作出整体性观察，需要采用专门的方法，以表或图的形式进行整理和加工，将其内在的条理化结构直观地显示出来，以便于阅读和观察。对原始样本数据的这种整理和加工过程，在统计中被称做样本数据的描述。

此外，计算统计量的最终目的是要用于推断总体参数，但统计量的具体取值是根据样本中的全体数据计算得出的，这一计算过程本身又是对样本数据的一种概括和归纳，从这种意义上说，统计量的计算过程也可被看做是对样本数据的一种描述过程。

### 1.3 统计方法的基本构成

统计是一门方法科学，方法体系庞大而严谨，图 1-2 告诉我们，其全部方法可以明确地分解为三个构成部分，即抽样方法、描述方法和推断方法。

#### 1.3.1 抽样方法

从总体中抽取样本，这是统计研究的起点。抽样方法多种多样，其主要构成内容如图 1-3 所示。

为保证推断结果的准确性和可靠性，样本的抽取需要采用专门的方法。抽样方法大体分为两类：随机抽样和非随机抽样。建立在概率论与数理统计基础上的抽样方法应当是随机抽样，这类方法可以保证最终的推断结果具备确定的把握程度和准确程度。非随机抽样也不失为一类可行的抽样方法，操作简便，易于掌握，但其推断结果不具备确定的把握程度和准确程度。

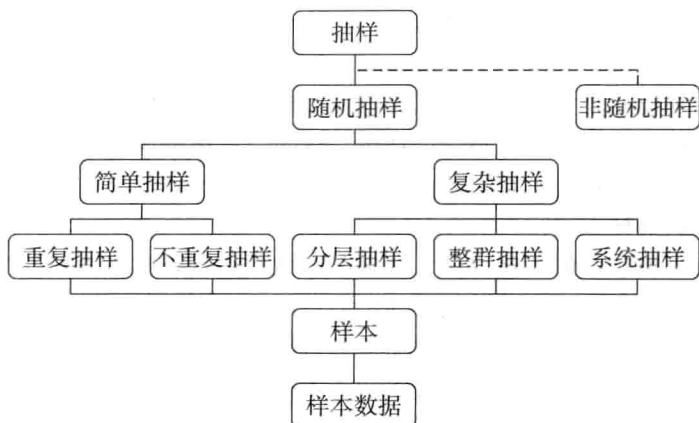


图 1-3 抽样方法的主要构成内容

### 1) 简单随机抽样

针对个体构成比较单纯划一的总体，进行小规模抽样，可采用简单随机抽样。这是实践中应用最为广泛的一种抽样方法。简单随机抽样具体又分为重复抽样与不重复抽样两种方式。

从总体中随机抽取一个个体获取观测值之后，就把这个个体放回总体中，再从总体中随机抽取第二个个体，依此类推，直至抽取了  $n$  个个体，从而形成一个容量为  $n$  的样本为止。在此过程中，总体中的每一个个体都有被抽中的可能，而且有可能被重复抽中，所以把这种抽样方法称做重复抽样。

不重复抽样是指从总体中随机抽取一个个体获取观测值之后，不再放回总体中，再从总体中随机抽取第二个个体，直至抽取了  $n$  个个体，从而形成一个容量为  $n$  的样本为止。在此过程中，总体中的每一个个体都有被抽中的可能，但不会有被重复抽中的情况，所以把这种抽样方法称做不重复抽样。

自有限总体的重复抽样和自无限总体的不重复抽样是最为典型的简单随机抽样，它们有一个共同的特点，即样本是在总体成分始终保持不变的过程中形成的。譬如，自容量为  $N$  的总体中抽取容量为  $n$  的样本，采取重复抽样时，总体容量始终为  $N$ ，成分是不变的。再譬如，自无限总体中抽取容量为  $n$  的样本，采取不重复抽样时，总体容量始终为无限大，成分也是不变的。自有限总体的重复抽样和自无限总体的不重复抽样在统计中又被称做纯随机抽样。纯随机抽样方法原理是形成其余较为复杂一些抽样方法的基础。本教材中有关统计推断方法原理的讲解，主要是建立在纯随机抽样的基础上的。

实践中，特别是在社会经济统计活动中，我们经常遇到的总体多为有限总体。需要注意的是，有限总体的不重复抽样与纯随机抽样是不同的，它是在总体成分不断改变的情况下形成样本的，其样本结构以及此后的估计和运算都与纯随机抽样有所区别。但由于其抽样操作过程比重复抽样要简便，所以在实践活动中比较常用。通常情况下为计算上的简便，只要总体容量充分大，也可以将这种抽样视同纯随机

抽样, 把握的标准一般为  $n/N < 0.05$ 。

## 2) 分层抽样

对于那些规模比较大、个体构成比较复杂的总体, 实施简单随机抽样往往会大量增加调查成本, 而且效果也不一定理想。这时, 可根据统计研究的目的以及相关的限制条件, 在不破坏随机原则的前提下, 选用分层抽样、整群抽样或系统抽样等较为复杂一些的抽样方法。

如果总体中所包含的大量个体, 分别归于性质上具有明显差异的不同类别, 出于调查成本或保持样本具有充分代表性上的考虑, 可将这些不同的类别看做是总体中的子总体, 先从各个子总体中按照简单随机抽样的方法抽取一定数量的个体, 再将这些被抽中的个体合并起来形成最终的样本, 这种抽样方法称做分层抽样, 又称分类抽样。

分层抽样做得好, 不仅可以节省调查成本, 而且会大大提高样本对总体的代表性, 还可以用来对不同子总体进行独立分析。

## 3) 整群抽样

如果总体容量过大, 出于大幅度降低调查成本的考虑, 可先将总体中的全部个体区分为容量及结构相似的不同的子总体, 以子总体为单位按照简单随机抽样的方法抽取一定数量的子总体, 再将这些被抽中的子总体中所包含的个体合并起来形成最终的样本, 这种抽样方法称做整群抽样。

整群抽样最明显的优点就是实施便利、成本低。但以群来划分子总体时, 要注意保证群内个体之间的差异性以及每一个群对总体的代表性, 理想的情况是每一个群都是总体的一个缩影, 否则容易造成较大误差。

## 4) 系统抽样

如果所需的样本容量很大, 样本中每一个个体的抽取过程就会变得非常繁琐, 为便利起见, 可对总体中的每一个个体进行编号, 并顺序地排成一列, 再随机地确定一个号码为抽样起点, 然后, 每隔一定的距离抽取一个个体, 直至抽取了一个容量为  $n$  的样本为止, 这种抽样方法称做系统抽样, 又称等距抽样。

在系统抽样中, 抽样起点一经确定, 其余的人样个体也就随之确定, 实施起来非常便利, 实践中常用来替代简单随机抽样。而且, 在总体中的个体之间存在较大差异的情况下, 系统抽样所获取的样本往往比简单随机抽样更具有代表性。

### 1.3.2 描述方法

样本生成于总体, 采取任何一种抽样方法所获取的样本数据, 其中都包含着有关总体的重要信息。认识和了解总体要从认识和了解样本数据开始。

然而, 原始的样本数据往往是大量而零散的, 根本就无法直接对其进行整体性观察, 其内在的结构条理必须借助某种专门的方法加以整理和概括, 才能够直观而准确地被描述出来。对样本数据的描述方法可分为两大类: 图表描述与统计量描述, 具体内容如图 1-4 所示。