数据处理 ■ 统计 ■ 分析

Excel 数据统计与分析范例应用

杨世莹 / 编著

本书由台湾著名Excel专家精心编著，以范例的方式向读者详细介绍如何使用 Excel 中的函数和分析工具来统计、处理数据，是追求 Excel 更上层楼的绝佳选择

本书适用于 Excel 2000/2002/2003 多个版本，是 Excel 软件使用者、数据统计分析人员、公司办公人员，以及相关专业学生的必备参考书

本书特色：

- ◆专门介绍了 Excel 数据统计基础和应用范例
- ◆所举数据和范例均出自统计工作的第一手资料
- ◆多达 10 个大类，近 100 种数据统计分析方法
- ◆完全按照统计学知识结构划分章节、安排范例
- ◆提供丰富多样的课后习题，帮助读者学习参考
- ◆非常适合作为统计学及相关专业的教学用书

免费附赠本书所有实例的原始数据、源文件及课后练习文档，下载请访问 www.21books.com

中国青年出版社
中国青年电子出版社
http://www.21books.com http://www.cgchina.com

旗标出版股份有限公司

数据处理■统计■分析

Excel 数据统计与分析范例应用

杨世莹/编著

中国青年出版社
中国青年电子出版社
http://www.21books.com http://www.cgchina.com

旗标出版股份有限公司

（京）新登字 083 号

图书在版编目(CIP)数据

Excel 数据统计与分析范例应用 / 杨世莹编著. —北京：中国青年出版社，2004

ISBN 7-5006-5687-4

I.E... II.杨... III.电子表格系统，Excel —应用软件—统计分析 IV. C819

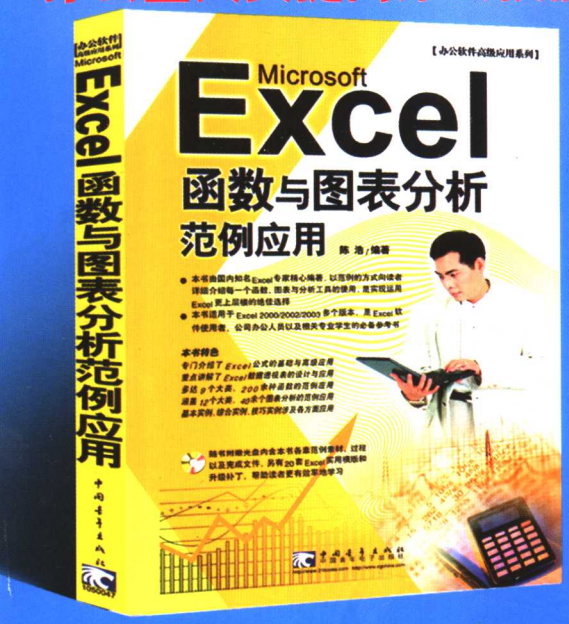
中国版本图书馆 CIP 数据核字（2004）第 118039 号

书　　名：Excel 数据统计与分析范例应用
编　　著：杨世莹
出版发行：中国青年出版社
地址：北京市东四十二条 21 号　邮政编码：100708
电话：(010) 84015588　传真：(010) 64053266
印　　刷：山东高唐印刷责任有限公司
开　　本：787 × 1092　1/16　印　　张：27
版　　次：2005 年 1 月北京第 1 版
印　　次：2005 年 1 月第 1 次印刷
书　　号：ISBN 7-5006-5687-4/TP・417
定　　价：39.00 元

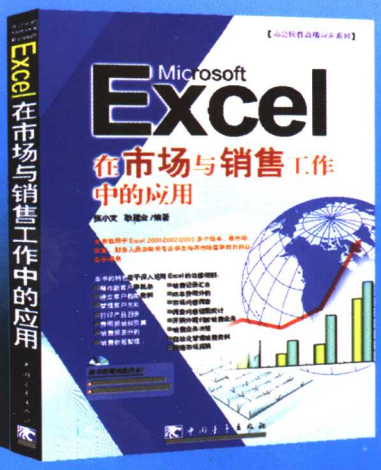
Excel 软件高效办公系列

中国青年出版社
中国青年电子出版社

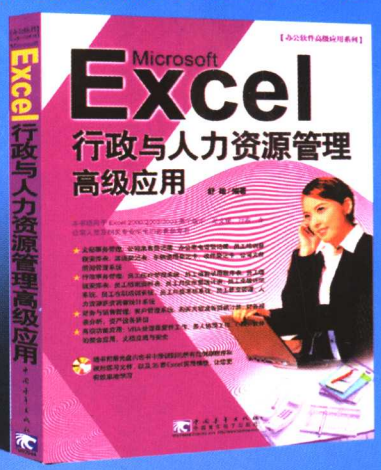
Excel 入门者提高应用水平的最佳选择
办公室人员提高办公技能的案头宝典



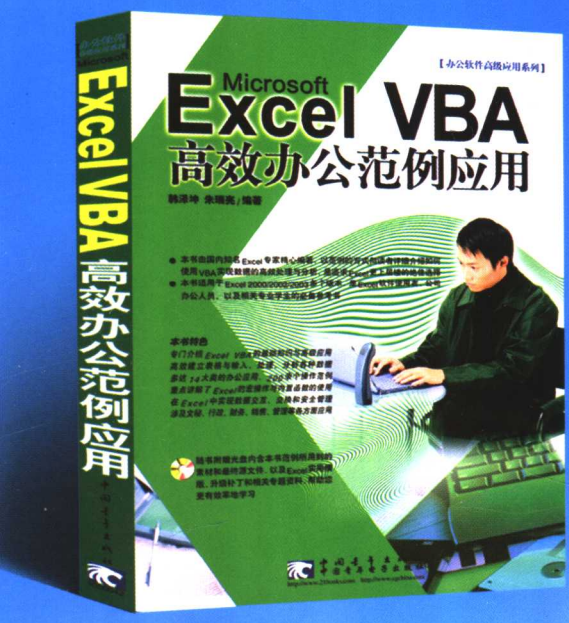
16开 / 410页 / 黑白 /1CD/ 39.00元



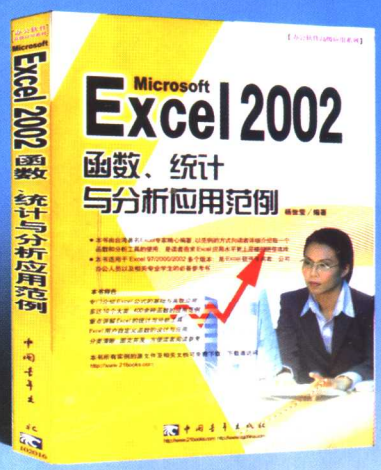
16开 / 381页 / 黑白 /1CD/ 38.00元



16开 / 388页 / 黑白 /1CD/ 39.00元



16开 / 415页 / 黑白 /1CD/39.00元



16开 / 385页 / 黑白 / 39.80元

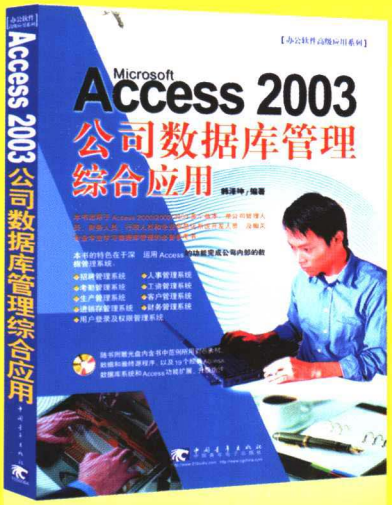
地址：北京东城区东四十条94号万信商务大厦502室 邮编：100007
电话：010-84015588 传真：010-64053266

办公软件高级应用系列

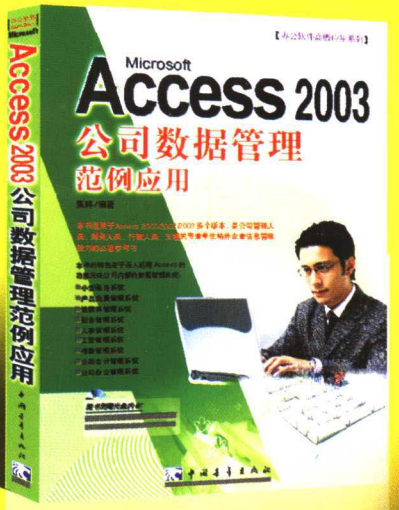
最新办公软件【经典畅销书】

平面设计 / 网页设计 / 3D动画 / 视频特技 / 室内设计 / 建筑设计 / 工业设计 / 网络管理 / 电脑入门 / 办公应用 / 高校教材

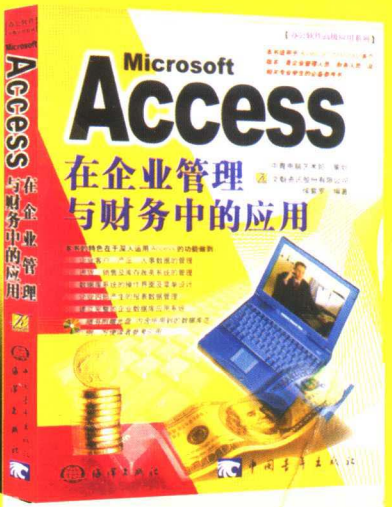
★ 2003年持续畅销全国各大书城及计算机专业书店
★ 公司白领提高办公技能的最佳案头宝典
★ 公司岗位培训必备教材



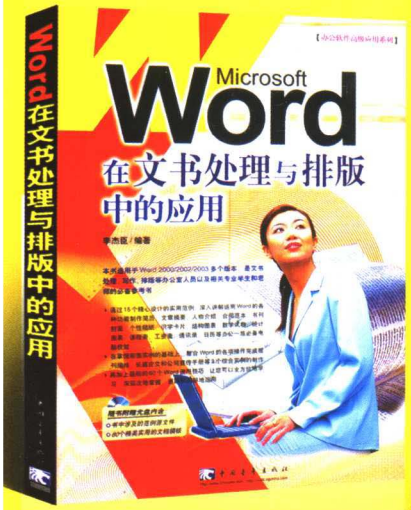
16开 / 429页 / 黑白 /
1CD/ 39.00元



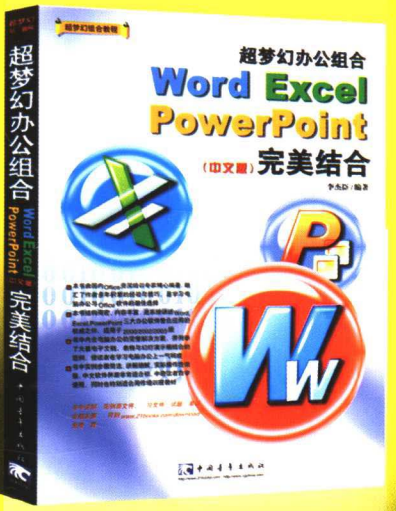
16开 / 397页 / 黑白 /
1CD/ 39.00元



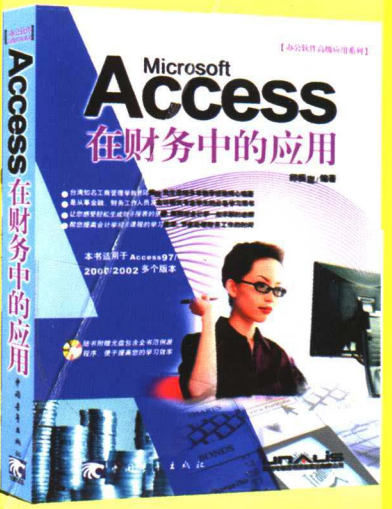
16开 / 412页 / 黑白 /
1CD/ 39.80元



16开 / 321页 / 黑白 /
1CD/ 32.00元



16开 / 328页 / 黑白 /
32.00元



16开 / 546页 / 黑白 /
1CD/ 48.00元

中国青年出版社 | 地址：北京东城区东四十条94号万信商务大厦502室
邮编 100007 电话：010-84015588 传真：010-64053266
http://www.21books.com http://www.cgchina.com

前言

虽然 Minitab, SPSS, SAS 等知名的统计分析软件的计算能力超强、提供的统计结果非常详尽，但它们在市面上的普及率却非常低，在个人电脑上的安装率甚至连全球的千分之一都不到。若利用这类高级统计分析软件来学习统计技巧，将来离开学校后会很容易面临无适当软件可用的窘境。纵有一身绝技也难以发挥。

由于微软的 Office 已相当普及，而且广泛地被工商企业及个人使用。要想在一台个人电脑上找到 Excel 软件，要比找 SPSS 或 SAS 软件容易多了，且其具有易学易懂的特性，所以本书决定以 Excel 为工具来学习统计技巧。虽然 Excel 并非被归类为统计软件，且它在统计方面的功能是绝对无法与 SPSS 或 SAS 相提并论的，但对绝大多数人而言，Excel 在统计方面的应用已经足够用了。

本书由台湾著名 Excel 专家精心编著，以范例的方式向读者详细介绍如何使用 Excel 中的函数和分析工具来统计、处理数据。本书以笔者多年在教学过程中进行实际问卷调查所收集到的资料作为全书分析实例的主要资料，其特性为：

1. 覆盖面广

完全按照统计学知识结构划分章节和安排范例。

2. 结构合理

内容涵盖频率分布交叉分析、集中趋势、离散程度、假设检验、单因子方差分析相关性、回归性等 10 大类近 100 种 Excel 数据统计和分析方法。

3. 实例充足

笔者从教多年收集整理了丰富的原始问卷和资料，且所举数据和范例均出自统计工作的第一手资料。

4. 过程和说明并重

每一个实例，除了不厌其烦、详尽地解说其操作及运算过程外，对其结果也尽量以读者较容易接受的口语方式进行阐释，而不是用难懂的统计术语讲解。

5. 广泛适用

本书适用于 Excel 2000/2002/2003 多个版本，是 Excel 软件使用者、数据统计分析人员、公司办公人员，以及相关专业学生的必备参考书，也非常适合作为统计学及相关专业的教学用书。

为了让学习者节省时间，并有更多的自我练习的机会，每一范例均再加入一个含有题目内容的练习工作表，而且在章节适当位置还加有“练习”的题目，各章章末还备有习题。将每一章所使用的范例、“练习”与习题题目均附于随书附送文件中，读者可以随时将某一章插进来阅读并练习。（本书每一范例、练习或习题，均将画面缩小到可以看到的底下工作表标签的程度，最主要的是方便读者在所附送的文件中，找到正确的工作表）。

撰写本书虽力求结构完整与内容详尽，但仍有疏漏和错误，诚盼各界人士与读者予以指正。

作者　杨世莹

2005 年 1 月

目录 contents

第 1 章 概述

第 2 章 研究程序与抽样

第 3 章 样本大小

第 4 章 设计问卷与取得数据

第 5 章 频率分布

第 6 章 交叉分析表

contents

第 7 章 集中趋势

第 8 章 离散程度

第 9 章 估计

第 10 章 假设检验

第 11 章 单因子方差分析

第 12 章 相关

目录

contents

第13章 回归

概　　述

Chapter 1

Excel 2003

1-1 传统的统计学教法

对于很多学过或正在学习统计学的人来说，统计一直是心中永远的痛。课堂上，传统的统计学是讲解一大串的定义、定理与公式，除了推导公式外还要加以证明。最后还得在不使用计算机的情况下，以笔算计算出各种统计结果。

这些定义、定理与公式，学生通常无法完全理解。但为了考试，只好将其死记硬背下来。目的没有别的，只求及格过关而已。等考完试，所记忆的公式几乎完全忘光，一点印象也没留下。在这种情况下的学习，想要学生能真正学好、做到融会贯通，实在是不可能。更别说要他们学后离开学校，还能学以致用了。

1-2 新式的统计学教法

最近国内外很多高等院校中的新一代统计人才，开始使用 Minitab, SPSS (Statistical Package for the Social Science) , SAS (Statistical Analysis Software) 等统计分析软件来配合教学。

这些软件的计算能力超强，提供的统计结果非常详尽，所有书本上或不在书本上的计算方法与统计量均可计算出来。学生不必记忆各项计算方法及公式，只须知道如何将数据输入即可，过去还得透过简单的公式，现在则只需直接点击鼠标选取命令和选项，即可进行分析，从而获得所要的统计数字。

从此学生终于可以脱离痛苦的深渊。可以不必将所有精神花费在记忆公式上，可以将时间用在如何对所获得的统计数字进行描述、分析、解释，用来了解其意义，进而加以应用到所学的各领域上 (例如企业管理、工业管理、政治、经济、社会、心理学、医学、生物、法律、农业等)。

1-3 为何要使用 Excel 来学习统计

前面介绍的 Minitab, SPSS, SAS 等统计分析软件，在市面上的普及率非常的低，在个人电脑上，占有率甚至连千分之一都不到。其原因如下：

1. 价格昂贵

价位高达上万到数十万。通常，教育单位、国家机关或大型研究单位才有能力购买。一般学生、社会人士或中小企业公司都因负担不起而不可能购买。

2. 学习困难

这几个软件，若仅是要进行操作而取得分析结果，在有人指导下或有相关优秀书籍来参考的情况下并不困难。但因学习的人不多，受过完整训练的师资本来就不多，市面上可用的书籍也不是很多。所以，对大部分人来说，还是很困难。况且很多单位所使用的还是英文版，更加大了学习上的困难。

事实上SAS软件较适合学习统计理论时使用，无论操作或编写程序均较为困难。而SPSS软件则较适合分析市场调查的数据，通常不需要编写程序，即可进行操作。

3. 报表难懂

这几个高级的统计软件，所分析出来的统计数字相当多，很多是读者学统计时从来就没看过的统计值，根本就不知道其作用。而事实上，一般统计应用所使用的统计数字并不很复杂，主要目的是要让大部分人能看得懂。计算并列出这些无法拿来应用的统计数字，不只是浪费资源，更是对学习者信心的一大打击。

若统计数据只有少数几个人能看懂，其作用将大打折扣。例如，市场调查的分析结果，如果只有少数几个研究人员能懂，老板或主管又怎能有信心根据一个完全不懂的结果来做决策？又如政策支持度的结果，只要简单的几个百分比大概也就够了，列出一大串的相关统计值，不仅管理者看不懂，各报章杂志的记者及主编也看不懂，刊登出来后，读者也看不懂，如何会引起共鸣？政策又将如何修正或推行？

若以这类高级统计分析软件来学习统计，因普及率过低，将来离开学校后很容易会面临到没有适当软件可用的窘境。纵有一身绝技，也难以发挥。

由于微软的Office已相当普及，并且广泛地应用于工商企业及个人使用领域。要想在一部个人电脑上找到Excel，要比找到SPSS或SAS软件容易多了，而且Excel具有易学易懂的特性。所以本书决定以Excel为工具来帮助读者学习统计。虽然Excel并没有被归类为统计软件，并且其与统计有关的命令、函数或命令集的功能是绝对无法与SPSS或SAS软件相提并论的，但对绝大多数人而言已经足够了。

1-4 本书的特色

本书是将作者在教学时，经多年与学生合作，进行实际问卷调查所收集到的数据，作为全书分析实例的主要数据。其特性为：

1. 覆盖面广

完全按照统计学知识结构划分章节和安排范例。

2. 结构合理

内容涵盖频率分布交叉分析、集中趋势、离散程度、假设检验、单因子方差分析相关性、回归性等10大类近100种Excel数据统计和分析方法。

3. 实例充足

笔者从教多年收集整理了丰富的原始问卷和资料，且所举数据和范例均出自统计工作的第一手资料。

4. 过程和说明并重

每一个实例，除了不厌其烦、详尽地解说其操作及运算过程外，对其结果也尽量以读者较容易接受的口语方式进行阐释，而不是用难懂的统计术语讲解。

5. 广泛适用

本书适用于 Excel 2000/2002/2003 多个版本，是 Excel 软件使用者、数据统计分析人员、公司办公人员，以及相关专业学生的必备参考书，也非常适合作为统计学及相关专业的教学用书。

6. 易学易用

所讲到的内容均是一般常用的统计技巧。既没有导出、验证公式等繁琐的内容，也没有过深的理论。绝对能让读者“学得轻松、学以致用”。

事实上还有更多读者看不到的地方，也都正在使用统计技巧。如公司经营者进行市场调查，以此来预测各品牌的市场占有率或消费者对某一产品的偏好程度；银行通过来源于客户的基本数据，判断其信用程度，用来确定贷款或信用卡额度；医师对同一疾病的各种可用药物的疗效进行实验，以找出最适当的治疗法；工厂中产品检验工程师对产品抽样进行质量管理，用来确定一批产品是否合格；股民绘制股票交易图，以预测明天股价的走向等。

1-5 什么是统计

统计学 (Statistics) 是数学的一个分支，用以搜集、整理、分析数据，进而推导分析结果的科学方法，因而有学者也将统计学统称为统计方法 (Statistical Method)。统计不但用来简化和表示一组数据，而且主要是在探讨如何从一组数据的总体中（总体，Population），以某一抽取过程（抽样，Sampling）抽出部分数据（样本，Sample），来研究如何利用这一部分数据去估计、检验或预测数据总体的某些未知的特性值。其范围应包括：

1. 搜集数据

统计工作的第一步，即要搜集相关的统计数据，无论是原始数据（如已调查或实验所得的数据）还是二次数据（例如政府公布的统计数据）。统计数据的搜集必须切合统计分析的目的，否则将浪费许多时间和金钱，却得到一大堆无法使用的数据。

2. 整理数据

所搜集的数据通常是杂乱无章的，尤其是原始数据。所以需要加以整理，以计算出想要的统计数字或绘制出统计图表。此部分包括分类、归类、列表和绘图，通常是用电脑来加以整理。

3. 分析数据

所计算出的统计数字仍只是一大堆数字结果，并不是每个人都能够看得懂，必须由受过统计训练的专家来分析，才能解释出其涵义。例如，分析其集中趋势、方差、差异性、相关性、周期性等。此部分即描述性统计（Descriptive Statistics），主要是描述和叙述事物的现象。

4. 推论数据

如果分析对象只是一部分样本，则仍需要推论其分析结果，来推断总体的可能结果。例如，用样本统计量推算总体的参数（Parameter），或以样本的结果证实或否定一些有关总

体的假设。此部分即推论统计的工作，利用样本所收集的数据，推论总体的状况。

1-6 统计学的分类

以目前国内大学院校所开设的统计学课程通常可分为：

- 应用统计学（Applied Statistics）

着重于如何将统计方法应用到各种自然或社会科学上。其内容一般只涵盖基础统计学，由于统计的应用范围非常广泛，所以是绝大多数院校（农、工、商、文、理等）各科系的必修课程。

- 数理统计学（Mathematical Statistics）

探讨统计学的数学原理以及各统计方法的来源。通常是在修完基础统计学之后，才会继续选修的课程。大部分为主修统计的学生的必修课程。至于其他学科通常是将其列为选修，但也有较注重计量的极少数科系，会将其列为必修课程。

1-7 几个统计的专有名词

总体

总体是统计人员想要研究调查的所有对象，它是由一些具有某种共同性的基本单位所组成。若总体个数有限，其个数一般以大写的 N 表示。

总体可以是一群人，例如，想调查全省 12 岁以上的女性消费者的化妆品消费情况，其总体将为全省 12 岁以上的全体女性。总体也可以是一群事物，例如，汽车制造厂商想了解其所生产的某一款汽车引擎的故障情况，其总体即为该型汽车的所有引擎。

基本单位

基本单位是指总体中的个别元素，要根据抽样调查的目的来决定。如果调查目的是想了解品牌喜好的投票倾向，其基本单位将是具有投票权的每一个人。但若调查目的是想估计每户家庭中所拥有的个人电脑的数量，则其基本单位将是每一个家庭。

普查

普查 (census) 是对整个总体进行全面调查或研究。如果总体数目很大，那么进行普查所花费的时间、金钱以及动用的人员则都非常庞大。例如，统计局所进行的全国人口普查、户口及住宅普查、工商及服务业普查、农、林、渔、牧业普查等。

这几类大型普查通常除了国家或各级政府单位外，其他私人机构或组织是不太可能会实

施普查的，就算是国家也不可能经常进行此类大型的普查。

但如果总体数目并不大，则普查将是最能探讨出总体现象的一种调查方式。例如，只想知道某班 50 名学生此次旅游分别想要到何处去玩，对此进行一次普查将不是什么难事。

抽查

相对于普查的全面调查，抽查是在某一总体中抽出一小部分个体进行调查。

抽查的目的在于省时、省力、省钱。普查的结果虽然相对于抽查更精确，但往往费钱、费时而且费力，除了极为重要的调查外，一般很少使用普查的方式进行。

样本

样本是总体的一个部分，一个样本是由数个数值所组成，此数目通常以小写 n 表示，称为样本量或样本大小。

想知道整个总体的最正确的状况，当然是进行全面普查。但通常无法对总体进行全面普查，其主要原因是：

1. 经济

研究整个总体可能太花金钱，若总体数目很大，光印刷及邮寄问卷的花费就很可观；更不用说用人员来进行面访。同时，想获得整个总体全部的回卷，也几乎是不可能的。

2. 时效

无论对研究者还是对决策者而言，时效是非常重要的考虑因素。若一个研究因为要取得整个总体的数据而拖延时间，等研究结果出炉，其结论可能已不合时宜。对企业的决策者，将无法提供任何帮助。

3. 难以接触

事实上，可能根本无法找出整个总体的全体成员。例如，要针对全国 20 岁以上的所有人来调查是否赞成某一政策？谈何容易！这些人东跑西跑的分散在各地，有的在高山或海岛，有的人还不在国内，怎么可能将其全部找出来？理论上，虽然还是可以接触到，但其成本将难以估计。

4. 总体过大

很多研究对象的总体确实很大。像一些畅销全球的汽车、电器、日用品等，消费者数以万计或千万计，且散居全球各地，根本就不可能进行普查。

5. 毁坏性

若研究整个总体，每个商品都进行测试则可能会没有剩下的商品可以出售。例如，要测试轮胎的耐磨程度，将所生产的每一个轮胎都拿来进行路面实际驾驶或以机器来磨，以检查其是否合乎品质要求，磨完或驾驶过后的轮胎，就无法再送回到市面上销售了。

6. 正确性

普查费时、费力、费钱，到最后往往会草率了之，所获得的数据的正确性却并不是很

可靠。还不如针对较少数的抽样，进行仔细的数据搜集。所以，我们通常只能对总体进行抽样，抽出较少的几个样本进行分析。不过，我们真正想了解和研究的是总体，样本只是达到目的的手段。由了解和研究样本来描述或推论总体状况。虽然，样本所提供的信息量并不完整，但若经过妥善的抽样安排，其误差并不会太大。

统计学的目的就在于：利用样本中所获得的信息来推测总体的结果。所以可以说，统计学就是研究如何以最低成本的抽样调查或实验设计来搜集一定数量的信息，并将这些信息应用在推测总体上。

影响样本数大小的因素：

1. 总体大小：总体越大，所需样本越大。
2. 可用资源：可用资源越大 (金钱、时间、人力等)，可收集的样本就越大。
3. 可容忍的误差：可容忍的误差越小，所需样本越大，可以缩小误差。
4. 误差的代价：误差的代价 (产生误差的损失) 越大，所需样本要越大，可以减少损失。
5. 总体方差：总体方差越大，所需样本越大。总体的成员彼此间相似度很高，样本数就可以小一点；反之，总体的成员彼此间相似度很小，样本数就得大一点。

观察值

观测一个实验或统计问题，所记录下来的结果被称为观察值 (observation)，通常以小写 x 来表示。例如，为找出大学生每月零用钱的平均数，由每位受访者 (受测样本) 所提供的每月零用钱的数值就是一个观察值。

参数

参数 (parameters)，是总体的数值性叙述值，即用来描述总体某一特性的数字。例如，代表总体某一属性的数值：总体均值、总体方差、总体标准差和…。

参数通常用希腊字母来表示。例如，用 μ 表示总体均值、用 σ 表示总体标准差、用…表示所处理的数据是总体。

读者通常对总体的信息知道的不多，所以才需要由样本所获得的数据（统计值）来推论总体的参数值。

统计量

相对于参数，统计量（statistic）又称估计值（estimate），是样本的数值性叙述值，也就是用来描述样本某一特性的数字。例如，代表样本某一属性的数值：样本均值、样本标准差和…。这些值是根据样本所求得到，准备用来估计总体的参数值。

对于统计量，通常以英文字母来表示。例如：用 $\tilde{x}$ 或 $\bar{X}$ 表示样本均值、用 s^2 或 S^2 表示样本标准差、用…表示所处理的数据是样本而非总体。

抽样误差

抽样误差（sampling error）是指总体与样本之间的差异，由于样本并非总体，其间自然存有某些差异。其大小决定于以下两点：

1. 样本大小：样本越大，抽样误差将越小。除非普查，否则无法消除抽样误差。

2. 方差大小：方差是总体中各成员针对某一变量 (如年龄、所得等) 彼此间的差异，当方差越大，抽样误差将越大。

造成抽样误差的主要原因是抽样偏差（sampling bias），例如，仅在白天用电话对用户进行抽样调查，将漏掉白天不在家或家中没有电话的用户的意见，这就是一种抽样偏差，当然也会导致抽样误差。

变量

变量（variables）是用来描述总体中成员的某一特性。

在搜集数据的过程中，需要搜集各类的变量。例如，性别、年龄、职业、教育程度、收入等人口统计变量。又如，为了预测明年的销售额，而所搜集到的数据有广告费、人事费、销售人员数等，这也是一种变量。

在现实生活中或自然界的一些现象，通常都不是单一变量可以描述得很清楚。例如，要描述某一个人，仅使用性别变量，说他或她是男性或是女性，肯定是无法说明白的。但随着增加变量，例如年龄、肤色、头发、身高、体重、种族等，可以逐渐描述得更清楚一些。

1-8 变量的分类

离散变量

离散变量（discrete variable）或称不连续变量、类别变量，例如，性别、使用的手机品牌、就读的班级、宗教信仰、参加的社团、喜好的运动、最常饮用的饮料类别、最喜欢的歌手、最喜欢的影星等，均属离散变量。

性别为男或女，只是描述性别的现象。将男性标示为 1 或将女性标示 2，仅是为了方便电脑处理，并没有任何大小或倍数的关系。直觉上读者可能认为 2>1，2 为 1 的两倍。但若转为口语化，将变为：女大于男，女为男的两倍。任谁都不可能同意。而且若其均值为 1.69，也不具任何意义。最多也只能知道此次调查的女性样本比男性样本多些而已。

连续变量

连续变量（continuous variable）。例如，成绩、年龄、收入、长度、距离、体重、