

MBA/MPA

系列教材
XILIE JIAOCAI

管理统计学

GUANLI
TONGJIXUE

第2版

缪柏其 万红燕 编著



中国科学技术大学出版社

管理统计学

(第2版)

缪柏其 万红燕 编著

中国科学技术大学出版社

内 容 简 介

本书以管理经济理论为背景,深入浅出地介绍了描述性统计方法和数理统计中关于均值的区间估计和假设检验中常用的统计方法,如正态检验、 t 检验、 χ 检验、Wilcoxon 秩和检验、符号检验和游程检验等.同时也介绍了有关相关与回归分析、时间序列分析.最后介绍了著名的 SPSS 统计软件、Excel 软件在统计中的应用.

本书可作为 MBA、MPA 学生以及工商管理研究生的教材,也可供有关管理人员阅读和参考.

图书在版编目(CIP)数据

管理统计学/缪柏其,万红燕编著. —2 版. —合肥:中国科学技术大学出版社, 2010. 2

(MBA/MPA 系列教材)

ISBN 978 - 7 - 312 - 02593 - 8

I. 管… II. ① 缪… ② 万… III. 经济统计学 IV. F222

中国版本图书馆 CIP 数据核字(2010)第 005015 号

中国科学技术大学出版社出版发行

地址 安徽省合肥市金寨路 96 号,230026

网址 <http://press.ustc.edu.cn>

合肥现代印务有限公司印刷

全国新华书店经销

开本:710×960 1/16 印张:20 字数:375 千

2002 年 9 月第 1 版 2010 年 2 月第 2 版

2010 年 2 月第 5 次印刷

印数:12001—15000 册

定价:33.00 元

第2版前言

统计学(包括描述性统计学和数理统计学)是一门研究不确定性的学科,它主要研究如何有效地收集、整理数据,并从这些数据中挖掘出有用的信息。目前,统计学不仅在工农业生产和生物学、物理学等学科的科学研究中得到广泛的应用,而且在地理环境科学、信息技术、社会人文和经济科学中得到广泛的应用。改革开放中的管理人员常常面临如何作出正确的科学决策的问题。所谓科学决策,就是凭数据决策而不是拍脑袋决策。当前我们面临的数据比以往任何时候都要多,并且结构也非常复杂,既有定性数据,如性别、职业,又有定量数据,如收入、产量;而且数据呈现空间性和时间性。因此在作出正确的决策之前,首先要学会如何获取对决策有用的数据。但是这些收集到的原始数据通常含有杂质并让人感到有些困惑。要使其具有易懂的解释,并能用于各种决策,就必须对原始数据进行归纳和整理,这就需要了解统计学。

比利时的数学家凯特勒(A. Quetelet, 1796~1874)最早把统计学应用于人类事务。经济学和人口统计学中的某些近代概念,如 GNP(国民生产总值)、增长率、发展率和人口增长率等都是凯特勒和他的弟子们提出的。

统计学是一门科学,它在建立软科学的经验规律中起着重要作用。作为量化研究和研究不确定性方法的统计学的基础和许多哲学观点有关,因此统计学能够对任一主题进行讨论。在这种意义下,统计学是关于一切学问的学问。著名的印度统计学家劳教授(C. R. Rao)在《统计与真理》一书的扉页上写道:

在终极的分析中,一切知识都是历史

在抽象的意义下,一切科学都是数学

在理性的基础上,所有的判断都是统计学

统计学是一种技术。统计学的方法论就是在工业生产中为了保证所需产品达到一定的质量标准和保持稳定性的管理系统中建立起来的,它能用于控制和减少不确定性,度量不确定性,极大发挥个人和政府的工作效率,从而作出科学决策。统

计学是探索真理必不可少的工具。

统计学也是一门艺术,对同一组数据,站在不同的立场,用不同的统计方法及不同的置信水平,可能得出不同的结论,因此如何正确使用统计方法依赖于使用者的统计技巧和经验。正是由于对统计学是艺术的理解不够,不少人对统计学产生了许多误解,如本杰明说过(美国作家马克·吐温引用过)“有三类谎言:谎言、可恶的谎言以及统计学”。事实上,应用统计学的原理和方法恰恰能够区分科学真理和科学虚伪。据国外报道,在美国许多人误用统计,在我国,统计学远远没有美国普及,因此误用统计的也一定不在少数。

作为一个管理者,其统计素质是个人综合素质中很重要的一个组成部分。我们正处于信息革命的时代,人们已经认识到信息的重要性,特别是健全的政策制定要依靠准确可靠的信息,而对信息的归纳利用有赖于统计知识。其次,每个人在一生中的重大问题上,如上学、结婚、投资及处理每天的工作,都必须作出各种决策,这就要求有统计推断的知识,这种知识有助于了解自然界和人类活动中的不确定性,在借鉴别人和利用自己经验作出决策时使风险最小化。“进一步,统计知识是个人的一笔财富,可以保护自己和家人不受传染病的影响,防范政治家的宣传和商人夸大事实的广告,摆脱掉比疾病还糟的迷信,有效地利用天气预报,了解各种灾害,如核电厂的放射线泄漏以及影响生活的其他自己不能控制的方面”(《统计与真理》第108面)。

由于在了解、解释信息时需要掌握一定的统计学知识,因此理解统计学对管理者及政策制定者尤为重要。对统计学的一知半解常常造成不必要的上当受骗,对统计学的一概排斥也往往造成不必要的愚昧无知。那么,统计学是不是很难掌握?我们的回答是否定的。本教材就是针对管理人员学习统计学的一本入门书,重点是让读者了解统计思想和统计概念,了解最基本的统计推断方法。只要从思想上真正认识到这门课程的重要性,就没有过不去的坎!

像今天有能力的公民能读会写一样,将来会有一天要求有能力的公民必须会计算,而且能够利用平均值、最大值和最小值,可以预期,这样的时代已经不远了(英国,威尔斯(H. G. Wells))。

当代社会已经进入信息社会和网络社会,管理者必须了解定量信息、数字的意义,学会收集、整理、分析和利用定量信息,懂得如何使用这些信息或数字,必须学会面对不确定性,而这一切都离不开统计方法的使用。因此现代管理者必须掌握一定的统计知识,具有一定的统计学素养。管理统计学是一门以一般统计学为工具研究社会和经济管理的应用学科。现在已广泛地应用于市场调查、证券市场分析、风险分析、产品抽样调查和质量控制、投资效益评估、人力资源分析和评估、经济指标

分析和预测、公共管理部门决策等方面,为管理者进行正确的决策提供了科学依据。

不知不觉7年过去了,在这7年中,这本教材4次重印。这次修订吸收了包括郑坚坚副教授等许多专家、同仁和读者的建议,在保持原书基本框架和写作风格的同时,每一章都或多或少进行了修改和补充。

在原书基本结构的基础上,万红燕副教授对前4章进行了大量的修改,使内容阐述上更加细致,所举的例子更加注重实用。考虑到MBA/MPA学员的知识特点,补充了Excel软件使用简介,描述了如何用Excel软件在数据处理、结果显示和计算手段方面对各种统计方法的实现。

总的来看,这一版突出增加了实用性内容,更加便于教师教学和学生学学习;同时改正了原先版本中的一些疏漏之处,使得本教材朝着更加完善、更加充实和更加实用的方向又进了一步。

每一次修订的过程,也是作者不断学习和不断提高的过程。欢迎广大读者、专家和同仁对本教材提出批评和建议。

缪柏其 万红燕
2009年6月18日
于中国科学技术大学

前 言

管理应用统计学是一门以经济管理理论为基础,以一般统计学为工具研究社会和经济管理的应用学科.它研究如何有效地收集、整理和解释反映社会和经济管理实践的数据,以期认识数据的规律性及内在的社会和经济涵义.管理统计学是应用统计的一部分,现在已广泛地应用于市场调查、证券市场分析、风险分析、产品抽样调查和质量控制、投资效益评估、人力资源分析和评估、经济指标分析和预测等方面,为管理者进行正确的决策提供了科学依据.以人为本的导向使管理方法进一步由定性向定量方向发展,而这一切都离不开统计方法的使用.因此一个现代管理者必须掌握一定的统计知识.

1. 课程的目的和要求

管理应用统计学是商学院的核心课程,本课程为 40 学时.课程的目的和要求是:

(i) 认识到管理统计学是一门科学,也是一门艺术.不同的问题要用不同的统计方法去解决.

(ii) 了解如何有效地收集、整理数据,并应用基本统计原理和方法来处理和分析数据.

(iii) 熟悉常用的统计方法和公式,并能正确地解释.

(iv) 了解并能初步应用常见的统计软件包(如 SPSS、Excel、SAS 等).

(v) 能正确地应用统计方法来研究和处理管理和经济问题中的数据.

2. 大纲的教学体系

管理统计学教学内容共 11 章,分为 6 个部分:

(1) 描述性统计和统计量

(i) 要点和要求

这一部分由前四章组成.首先要了解不是所有的数据都可以进行统计分析,我

们对数据有一定的要求. 因此有一个如何收集数据的问题. 在此基础上用列表、图示和加减乘除等初等方法对数据进行分析, 对数据表达的总体有直观的了解. 具体来说, 第 1 章是叙述如何有效地收集数据; 第 2 和第 3 章介绍一些常见的统计量; 第 4 章介绍统计指数的构成.

(ii) 预备知识

阅读一些统计报表、年鉴、商业和政府有关部门发表的数据.

(2) 统计推断的概率论基础

(i) 内容和要求

这部分由第 5 章和第 6 章组成. 为了对数据所反映的深层次的内涵进行分析, 我们需要一些概率论的知识. 描述性统计和数理统计的本质差别就在于后者用了概率论的知识. 我们要求学员能掌握离散分布中的二项分布和连续分布中的正态分布; 了解大数定理和中心极限定理.

(ii) 要点

概念: 样本空间、事件、独立性、二项分布、正态分布、均值、方差、相关系数;

公式: 全概率公式、贝叶斯公式、一般正态分布函数用标准正态分布函数表示的公式;

中心极限定理: 正态分布随机变量的线性组合仍是正态随机变量.

(3) 统计推断

(i) 内容和要求

这部分由第 7 章和第 8 章组成. 当我们按一定的规则从总体中抽取了一个样本后, 一个重要的问题是如何由样本来推断总体的性质. 在管理应用统计学中我们归纳为估计问题和检验问题. 在估计问题中, 我们假定总体的分布类型已知, 但其中的一些参数未知, 由样本对这些参数进行估计. 估计分为点估计和区间估计两类. 关于点估计, 我们最关心的是总体均值和方差的点估计. 在概率论的基础上, 以一定的置信度我们把总体均值和方差的真值估计在以点估计为基础的区间内. 在检验问题中, 我们分两种情况讨论. 一种是总体类型已知, 对其中的参数进行统计检验. 总体均值的区间估计和参数检验依赖于正态性, 在大样本情况下可以由中心极限定理来近似. 另一种情况是总体的分布类型未知, 我们需要对其分布进行检验, 这称为非参数检验. 一个重要的非参数检验统计量是 χ^2 统计量, 它也常用来检验列联表中两个变量是否独立.

在这一部分中我们要求学员掌握方差已知和未知情况下总体均值和方差的点估计、均值的区间估计和检验、列联表检验和拟合优度检验.

(ii) 要点

概念:点估计、置信度、精度、自由度、区间估计、假设检验、成组比较、成对比较、列联表检验、拟合优度检验;

检验统计量: u 检验(标准正态分布)、 t 检验(t 分布)、 χ^2 检验(χ^2 统计量)。

(4) 相关分析和回归分析

(i) 内容和要求

本部分内容由第 9 章组成. 在相关分析中主要介绍有序数据和比例数据相关系数的概念; 在回归分析中我们着重介绍了一元线性回归系数的最小二乘估计方法以及有关的预测问题, 同时我们也讨论了用变量代换可以化为一元线性回归的非线性回归问题. 多元线性回归没作过多的介绍, 因为除了在技术上要用矩阵外, 估计和预测方法和一元线性回归思路是一样的. 当然, 多元线性回归中的变量选择是多元线性回归中特有的, 但已超出了本书的范围. 我们的要求是掌握相关系数的概念以及一元线性回归系数的估计方法和有关的预测问题.

(ii) 要点

概念: 相关系数、Spearman 相关系数、散点图、回归方程、回归直线、最小二乘估计、预测。

(5) 时间序列分析

(i) 内容和要求

在第 10 章我们讨论时间序列的分析. 关于时间序列的分析有两种方法: 一种是描述性的; 另一种是用统计知识建模的. 根据对 MBA 学员的要求, 我们重点介绍描述性的分析方法. 用统计知识建模的时间序列我们仅仅介绍一些概念以方便大家使用有关的统计软件包. 时间序列的描述性分析方法包括加法模型、乘法模型和滑动平均模型. 要求学员能对时间序列进行分解, 对季节因子进行修整, 并能应用所学时间序列模型进行预测。

(ii) 要点

概念: 趋势、季节因子、循环因子、误差因子、指数加权滑动平均、预测;

模型: 加法模型、乘法模型、滑动平均模型、自回归模型、自回归滑动平均混合模型、季节模型。

(6) 统计软件包

(i) 内容

本章介绍如何使用 SPSS 统计软件包. 在本章中我们从最基本的输入数据开

始,一步步地介绍本书前十章中的统计方法如何具体地在计算机上实施.

(ii) 要求

通过上机实习,初步学会使用 SPSS 统计软件包来分析和处理数据.

根据 MBA 大纲要求,我们编写了该书.其中第 1 章至第 10 章由缪柏其教授编著,第 11 章由蒋涛博士、彭衡博士和陈文志博士编著.陈文志博士绘制了包括附表中的所有图表.成书之前的讲义在中国科学技术大学 MBA 中心、管理科学与工程专业研究生课程班和安徽工商管理学院作为管理应用统计学教材各讲授了很多次.在讲授过程中得到了诸多同仁的关心和支持.吴耀华教授、汪道明教授、杨桂元教授、杜雪樵教授和郑坚坚副教授仔细审阅了本书的讲义,提出了不少有益的建议,并指出了许多笔误.在本书的写作过程中也得到了商学院领导方兆本教授、梁樑教授和商学院 MBA 中心储雪林教授和王苏鸣高级工程师等的大力支持.同时得到我妻子曹蓓和女儿的大力支持.在此谨向他们表示衷心的感谢!

缪柏其

2002 年 4 月 18 日

于中国科学技术大学

目 次

第 2 版前言	(i)
前 言	(v)
第 1 章 数据的收集	(1)
1.1 数据收集导论	(1)
1.2 数据来源	(4)
1.3 问卷设计	(10)
1.4 数据的偏差	(14)
1.5 数据的类型	(17)
第 2 章 数据的表示	(19)
2.1 数据的列表表示	(19)
2.2 数据的图形表示	(27)
第 3 章 描述性统计量	(36)
3.1 刻画数据集中程度的特征量	(36)
3.2 刻画数据离散程度的特征量	(46)
3.3 分布的偏度和峰度指标	(50)
第 4 章 统计指数	(55)
4.1 统计指数的概念	(55)
4.2 总指数的编制及应用	(60)
4.3 居民消费价格指数	(66)
第 5 章 随机试验、随机事件和概率	(71)
5.1 随机事件	(71)
5.2 概率的定义	(74)

5.3 事件的独立性	(77)
5.4 非独立事件与运算	(78)
第6章 随机变量及其概率分布	(88)
6.1 离散型随机变量的概率分布	(89)
6.2 连续型随机变量的概率分布	(93)
6.3 随机变量的数字特征	(99)
6.4 独立随机变量和(差)的均值和方差	(101)
6.5 中心极限定理	(102)
第7章 参数统计推断	(106)
7.1 什么是统计推断	(106)
7.2 总体均值的估计	(107)
7.3 总体百分比(数)和两总体百分比(数)差的估计	(117)
7.4 中位数的区间估计	(122)
7.5 单总体参数的假设检验	(123)
7.6 两类错误	(135)
7.7 两总体参数差的假设检验问题	(137)
第8章 非参数检验	(149)
8.1 χ^2 检验	(150)
8.2 威尔柯克逊秩和检验	(157)
8.3 符号检验	(161)
8.4 威尔柯克逊符号秩和检验	(163)
8.5 游程检验	(167)
第9章 相关分析和回归分析	(171)
9.1 相关关系	(171)
9.2 相关分析	(172)
9.3 一元线性回归分析	(179)
9.4 $\hat{a}, \hat{b}$ 的检验和预测	(186)
9.5 可化为线性函数的非线性回归	(189)
9.6 多元线性回归分析	(193)
第10章 时间序列分析	(196)
10.1 因子与模型	(198)

10.2	长期趋势的测定	(200)
10.3	季节因子的测定	(205)
10.4	循环因子	(209)
10.5	随机因素和残差	(210)
10.6	预报	(212)
10.7	指数加权滑动平均	(213)
10.8	统计时间序列模型简介	(215)
第 11 章	SPSS 统计软件使用简介	(220)
11.1	SPSS 简介	(220)
11.2	SPSS 的基本操作简介	(220)
11.3	SPSS 的统计分析示例	(236)
第 12 章	Excel 软件使用简介	(250)
12.1	Excel 简介	(250)
12.2	Excel 的数据处理	(256)
12.3	Excel 的统计分析示例	(265)
附表 1	累积二项分布表	(276)
附表 2	泊松分布表	(288)
附表 3	标准正态分布表	(290)
附表 4	标准正态分布双侧上分位点 $u_{\alpha/2}$ 表	(292)
附表 5	t 分布上 α 分位点表	(293)
附表 6	χ^2 分布上 α 分位点表	(295)
附表 7	威尔柯克逊秩和检验临界值表($\alpha = 0.01$)	(297)
附表 8	威尔柯克逊秩和检验临界值表($\alpha = 0.025$)	(298)
附表 9	威尔柯克逊秩和检验临界值表($\alpha = 0.05$)	(299)
附表 10	威尔柯克逊符号秩和检验临界值表	(300)
附表 11	游程检验表	(301)
附表 12	Durbin Watson 序列相关检验表($\alpha = 0.05$)	(303)
参考文献		(305)

第 1 章 数据的收集

1.1 数据收集导论

1. 引言

统计这个名词大家一定不陌生,它涉及人类生活的各个领域.例如,在计划户外活动时,我们需要得到约一周时间的详细天气预报;各种股票市场价格的信息可以告诉我们什么样的投资可以获利.

这里出现的一些数字通常是由统计得到的.统计就是指这样一些由数字表示的事件.当然,统计的内容远比这些数字的计算和获取多得多.

广义而言,统计可以表示对数据资料进行收集、整理和分析的活动过程,或者是对其活动过程所取得的成果体现,也可以是关于数据资料收集、整理、分析和推断的科学.

统计学作为学术和研究的一个领域,历史相对比较短,但作为数值信息的历史却很悠久.在历史记载中,人们很早就开始收集资料,现保存下来的古代的资料中,有国家及其资源和国民的构成的数据信息,说明统计学(Statistics)一词的词根来自国家(State)这一事实.时至今日,统计仍然是世界上各个层次的政府机构的支柱.现在统计的含义已经扩展到包括任何资料的收集,例如工业统计、人口统计等.除来自国家的起源外,统计学还有两个重要意思,一是统计可以认为是某种形式的数据,统计在语源学中的定义在某种意义上是“数据”,如某地的工业增加值、国家债务的规模等.现代社会对数据有一种永不满足的需求,正是基于这种需求,人们收集越来越多的数据,数据本身不是问题的解答,但是对我们以什么样的程度来回答案,以及在一个特殊的答案中含有多大程度的不确定性,或者对答案的信赖程度有多大这些问题的考虑来说,数据是基本的资料.二是人们更感兴趣的是如何收

集、处理数据,并通过统计学的帮助把数据中的信息变成有用的知识。

统计学是一门运用概率论和其他数学的方法和技术,用以研究社会、经济与自然科学的通用性方法论学科。统计学分为描述性统计学和数理统计学两大类。描述性统计学指的是如何从已知的观察资料,收集、整理、分析研究并提供统计资料的理论和方法。描述统计包括各种数据处理,这些数据的处理被用来总结或描述数据的某些重要特征,但不必深入地去试图论及数据中包含的深层次的信息。因此描述统计的主要作用是通过对现象进行调查或观察,然后将所得到的大量数据加以整理、简缩,制成统计图表,并就这些数据的分布特征(如集中趋势、离散趋势等)计算出一些反映数据某种特点的数字(如平均数、中位数、标准差等)。借助这些概括性数字,我们就可以从貌似杂乱无章的资料中提取有意义的信息,便于对不同的总体进行比较并作出结论。描述统计的基本思想是对某项研究中所收集的数据信息进行整理和表述。数理统计学是使用概率论和其他数量方法,研究如何有效地收集数据、分析数据以及进行统计推断的一门科学。简言之,数理统计学是一门以概率论为基础,研究如何由部分来推断整体某类特性的科学。借助于统计学的帮助,管理人员可以在一定的时间和范围内,从有效抽取的样本中对能否做某事作决策,从而达到节省时间、人力和物力的目的。可以说描述统计是数理统计的前提,数理统计是描述统计的深化。

广义而言,统计涉及如何收集和整理数据,如何对数据进行分析,如何对结论加以表述和解释等问题。收集、分析、表述、解释数据的主要目的是使管理者和决策制定者能更好地理解问题,获得更多的信息,作出更好的决策。

统计方法对于政府部门在政策的形成和评估上非常重要。政府为了收集和提供经济的、金融的、社会的和人口的统计数据并进行分析和预测,有一个庞大的管理机构。它们收集来的数据被用来制定日常行政工作以及各项计划中的政策。政策制定者在作出决策时,期望寻求技术指导,然而,重要的是他们自己在了解和解释信息时需要掌握某些专业技术知识,他们需要准确的信息,同时需要处理信息的准确的方法。同样,统计方法在很多专业领域如社会学、生物学、经济学、心理学等学科都作出了贡献。统计学为数据资料的使用提供了理论和方法的支持。

2. 数据收集的步骤

统计学是关于数据的科学,是一门能使人们成为熟练的数据生产者和高效率的数据使用者的科学,其内容包括数据的收集、分类、汇总、组织、分析和解释。

每个人都收集某种类型的数据,但商业上收集数据的量非常大,其中一些数据对商业决策是非常有用的。遗憾的是现实生活中虽然有许多公司在收集数据,但并

没有充分利用它们。

许多商业、经济和社会问题不能简单地用“是”和“不是”来回答。它们可能需要讨论及进一步明晰,也可能需要给出一个接受或拒绝的准则(Criteria)。为了考虑这些问题,需要对现有信息的完整性进行评估,然后提出还需要什么新信息。因此决策者不但需要数据,而且需要评估这些数据的质量。

例如,去年我国有多少垃圾被回收了?班级的大小会影响学生的学习成绩吗?人的健康和哪些因素有关?一种新产品的市场定位与哪些因素有关?等等。

为了回答这些以及其他的许多问题,必须收集相关的信息。在这些例子中,需要知道很多事情,表面上看,好像得到这些信息很容易,只需要询问或做一做试验即可。但是现在开始有问题了:谁去问?应该问谁?有足够的经费去问涉及该问题的每一个人吗?该问些什么?等等。在收集数据时我们需要考虑以下问题:

(1) 收集何类数据及收集的动机是什么?不考虑这个问题,则收集的数据可能范围太广,其中大部分会模糊你的真正目标甚至使研究陷于停顿。同样,为谁收集数据这个问题不仅决定收集内容,还决定数据处理及最后呈递数据的方式。

(2) 什么是我们考虑的相关总体,即调查对象。

(3) 数据从何处而来,即数据的来源问题。

(4) 分析所提问的人及如何被挑选出来的?如何从这些回答者中收集信息?

(5) 谁没有回答提问,即调查中的无反应问题。

(6) 收集的数据类型是什么?

此外,数据应有怎样的精确度界定了收集的方式、所需的期限及耗费的财力。许多收集来的数据精度不适当,或是精确度根本不明。而后一种情况由于令人无法弄清利用这些数据能干什么而会带来更多的问题。因此,必须进行某种形式的成本收益分析,因为成本必须明显低于潜在的收益。如果调查工作开销较大,其所带来的收益必须经仔细衡量。用同样的花费有可能购得更有效的研究材料。时间成本及其收益都必须考虑进去,因为决策者的信息需求适时而变,清晰明了的决策目标方案也有助于避免无用功的产生。

在数据收集中一个重要的概念是总体,即我们要分析的某一特定人群中某个指标取值的全体。因为数据收集是花钱、花时间、有代价的,因此首先要明确向哪些人群收集数据,例如你调查为什么人们买外国产的化妆品,但没有与购买国产化妆品的人接触,则你不可能区别如下事实:某些购买者没有意识到他们的化妆品是国外产的。所有新住宅房地产价格的集合和今年所有 MBA 或 MPA 毕业生起始年薪 5 万的集合等是总体的另外两个例子。

确定有关人群后,下一问题是知道这些人是谁(比如姓名、地址),称为抽样框

(Sampling Frame), 然后确定出调查内容。

什么是数据?一种描述性的定义是“可以由它作出推断的已知事情或事物”。数据一般是指从事统计或其他某种业务活动而产生的对事实的一种以量化、图表等形式出现的记录。这种记录可以被汇总或通过相应计算形成各种形式的数 据,并成为一种资源。这些资料和数据通过各种各类原始记录、账册、计算机和网络数据库、统计年鉴以及新闻媒体展现在社会和公众面前。数据有各种不同的表达形式,如可以是数字,还可能是类别、性别、学历等。

对决策者而言,数据的质量是一个最大的问题。首先统计数据特别是原始数据只是对事物或事件的一种量化记载,况且这种记录还有正确和错误之分,量化记录值和实际真值之间还有或多或少的误差存在。其次再假设这种记录恰当地反映了客观现象,其误差在允许范围之内的情况下,这些数据对管理和决策能否发挥信息的作用还取决于对数据的甄别、筛选、分类和整理的方法是否正确,角度是否恰当,因为数据有偏差或误导可以毁掉一个有效的决策过程,从虚假的数据出发,很难作出正确的决策。例如某个区行业的调查结果不能作为整个城市行业的结果。最后还需要考虑数据的完整性,决策者需要决定目前的数据对当前目标是否够用,或还需补充什么数据。资料的准确性对于决策是至关重要的。

1.2 数 据 来 源

数据的两大来源:原始数据(Primary Data)和二手数据(Secondary Data)。

原始数据来源如图 1.1 所示。

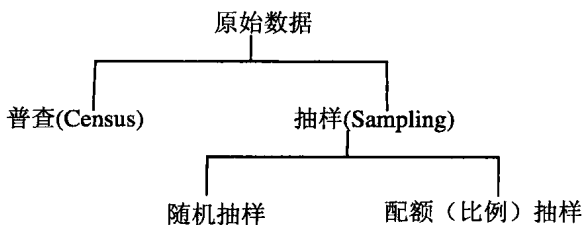


图 1.1 原始数据来源

原始数据:直接向调查对象收集所获得的数据。它的特点是为指定目的而