

普
华
经
管

Statistics

统计学

主编 江 岭 贾会远

副主编 朱维芳 梁 娟 李宗民 胡 源 赵 莉



人民邮电出版社
POSTS & TELECOM PRESS



Statistics

统计学

第 1 版

第 1 次印刷

1

统 计 学

主 编 江 岭 贾会远

副主编 朱维芳 梁 娟 李宗民

胡 源 赵 莉

人民邮电出版社

北 京

图书在版编目 (CIP) 数据

统计学 / 江岭, 贾会远主编. —北京: 人民邮电出版社, 2007.3

ISBN 978-7-115-15854-3

I. 统... II. ①江…②贾… III. 电子表格系统, Excel—应用—统计学 IV. C819

中国版本图书馆 CIP 数据核字 (2007) 第 022836 号

内 容 提 要

本书以统计理论为依据, 借鉴其他国内外相关教材和资料, 考虑到非统计专业的特点, 将统计学原理和统计实践有机地结合起来。本书内容分为三部分, 第一部分为统计学的基本理论, 包括统计基本概念、统计调查、统计整理、统计分布特征的描述、抽样推断、假设检验、方差分析、统计指数、时间数列等内容; 第二部分为国民经济统计分析; 第三部分为各种统计软件介绍。

本书可作为高等院校工商管理类学生使用, 也可作为统计爱好者及统计从业人员学习的参考书。

统计学

- ◆ 主 编 江 岭 贾会远
副 主 编 朱维芳 梁 娟 李宗民 胡 源 赵 莉
责任编辑 刘 盈
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京铭成印刷有限公司印刷
新华书店总店北京发行所经销
- ◆ 开本: 787×1092 1/16
印张: 21.25 2007 年 3 月第 1 版
字数: 375 千字 2007 年 3 月北京第 1 次印刷

ISBN 978-7-115-15854-3/F

定价: 29.80 元

读者服务热线: (010)67129879 印装质量热线: (010)67129223

前 言

《统计学》是为非统计专业本科生开设的一门基础必修课，设置本课程的目的在于培养学生有关统计知识方面的基本技能，培养学生应用统计方法分析问题和解决问题的实际能力。统计教学应达到的总体目标是：第一，使学生能系统地掌握各种统计方法，并理解各种统计方法中所包含的统计思想；第二，使学生掌握各种统计方法的不同特点、应用条件及适用场合；第三，培养学生运用统计方法分析问题和解决实际问题的能力。基于此，在过去的几年里，一方面，我们在教学中不断修改完善教案，更新了大部分习题和数据；另一方面，随着国内外最新的统计教材不断出版，使我们有了更多可供学习借鉴的同类优秀教材。正是有了这些基础，我们编写了此教材。

本书的特色主要如下。

1. 适用性强。本书在体系结构和内容安排上，不仅适用于经济学、工商管理类大学本、专科学生，也可以作为人文、社会科学以及其他学科的教材或参考书。

2. 易学性强。本书突出应用统计的特点，强调统计思想和应用，在充实内容和数据的同时，删去了数学推导和证明的内容，使得只有初等数学的读者就能理解、领会并掌握基本的统计分析思想和方法。

3. 整体性强。为了增强学生分析现实中复杂问题的能力，本书在详尽介绍统计理论与方法、推断统计方法的基础上，扩展了推断统计、统计分析等内容。

4. 实践性强。本书在分析统计理论的基础上，一些章节运用了 Excel 等统计分析工具，使理论与实践相结合，增强了理论的实践性和可操作性。

本书由江岭、贾会远任主编，朱维芳、梁娟、李宗民、胡源、赵莉任副主编。各章编写分工如下：第一、二章由赵莉编写，第三、十二章由朱维芳编写，第四、六章由江岭编写，第五章由潘前进编写，第七章由胡源编写，第八、九、十章由贾会远编写，第十一章和附录由梁娟编写。全书由江岭、贾会远、李宗民做最后修改定稿。

我们在编写过程中，参考了大量的教材和著作，在此向所有参考文献的作者和编者表示衷心的感谢。

由于作者水平所限，本书中仍有不少疏漏乃至错误之处，恳请同行和读者提出批评与建议。

编 者
2007 年

目 录

第一章 绪论	1
第一节 统计学的产生与发展	1
第二节 统计学的性质和特点	3
一、统计与统计学的涵义	3
二、统计规律与统计方法	6
三、统计学的性质和研究对象的特点	8
第三节 统计学的基本范畴	9
一、统计总体与样本	10
二、单位标志和标志表现	11
三、统计指标与指标体系	12
第四节 统计测定的层次	13
一、数据的计量尺度	13
二、数据类型	15
复习思考题	16
第二章 统计调查与整理	17
第一节 统计调查的组织形式	17
第二节 统计调查方案设计	19
一、统计数据的搜集方法	19
二、统计数据搜集方案	23
第三节 调查问卷设计	24
一、调查问卷的分类	25
二、问卷的基本结构	25
三、问卷的设计过程	26
第四节 统计分组和频数分布	28
一、统计分组	28
二、频数分布	33
第五节 统计图与统计表	36
一、统计表	36
二、统计图	39
三、统计分析报告	41
复习思考题	41
第三章 数据分布特征的测度	42

第一节 集中趋势的测度	42
一、定类数据集中趋势的测度——众数 (Mode)	42
二、定序数据集中趋势的测度——中位数和分位数	45
三、数值型数据集中趋势的测度	49
四、平均数之间的关系	58
第二节 离中趋势的测度	60
一、离中趋势概述	60
二、定类数据离散趋势的测度——异众比率	61
三、定序数据离散趋势的测度——四分位差	62
四、全距	62
五、平均差	63
六、方差和标准差	65
七、相对离散程度: 离散系数	70
第三节 偏态与峰度的测度	71
一、偏态的度量	72
二、峰度的度量	74
复习思考题	75
第四章 抽样推断	79
第一节 抽样推断的概念和作用	79
一、抽样推断的概念	79
二、抽样推断的特点	79
三、抽样推断的作用	79
四、抽样推断的理论基础	80
第二节 抽样推断的基本概念	81
一、全及总体和样本总体	81
二、全及指标和样本指标	82
三、样本容量与样本个数	82
四、重复抽样和不重复抽样	83
五、抽样框	83
第三节 抽样误差	84
一、抽样误差的概念	84
二、抽样平均误差	85
三、抽样极限误差	89
四、抽样误差的概率度	89
五、把握程度	90
第四节 点估计和区间估计	90
一、抽样估计的优良标准	90
二、点估计	91
三、区间估计	91

* 四、总体方差的区间估计	93
第五节 必要样本容量的确定	94
一、必要样本容量的计算公式	94
二、影响必要样本容量的因素	95
三、计算必要样本量应注意的问题	95
* 第六节 两总体参数的区间估计	96
一、两总体均值之差的区间估计	96
二、两总体比例之差的区间估计	97
三、估计两总体均值之差时样本容量的确定	98
四、估计两总体比例之差时样本容量的确定	98
五、两总体方差比的区间估计	99
第七节 抽样调查的组织方式及其推断	100
一、抽样推断的程序	100
二、简单随机抽样	101
三、分类抽样	102
四、等距抽样	106
五、整群抽样	108
六、多阶段抽样	110
七、抽样方式的选择	111
复习思考题	111
第五章 假设检验	114
第一节 假设检验的基本问题	114
一、假设检验的基本原理	114
二、统计假设与检验的三种基本形式	115
三、假设检验中的两类错误及其概率关系	116
四、假设检验步骤	118
第二节 正态总体均值、方差的假设检验	119
一、单个正态总体均值的检验	119
二、两个正态总体均值差的检验	121
第三节 单个总体方差的假设检验	123
一、总体均值已知时总体方差的检验	123
二、总体均值未知时总体方差的检验	124
复习思考题	125
第六章 方差分析	127
第一节 方差分析的基本问题	127
一、方差分析的基本术语	127
二、方差分析的基本原理	128
三、 F 分布	128

四、方差分析的基本条件	129
第二节 单因素方差分析	129
一、单因素方差分析的数据结构	129
二、单因素方差分析的步骤	129
三、关系强度的测定	133
四、Excel 进行单因素方差分析	133
第三节 双因素方差分析	135
一、双因素方差分析的类型	135
二、无交互作用的双因素方差分析	135
复习思考题	140
第七章 相关与回归分析	142
第一节 相关关系的概念与种类	142
一、相关关系的概念	142
二、相关关系的种类	142
三、相关关系的测定	143
第二节 一元线性回归分析	145
一、回归分析的概念	145
二、回归分析与相关分析	146
三、一元线性回归模型	146
四、估计标准差	149
五、回归方程的检验	151
六、回归预测	153
第三节 多元线性回归分析	154
一、多元线性回归模型	154
二、参数的 OLS 估计	154
三、Excel 应用举例	156
第四节 可线性化的曲线回归	157
可线性化的常用曲线类型	158
附录一 简单线性回归 Excel 上机实验	162
附录二 回归分析上机实验	164
复习思考题	165
第八章 时间数列分析	168
第一节 时间数列概述	168
一、时间数列的概念	168
二、时间数列的分类	168
三、编制时间数列的原则	171
四、时间数列常用分析方法	171
第二节 时间数列的水平指标	172

一、发展水平	172
二、平均发展水平	172
三、增长水平	179
四、平均增长水平	180
第三节 时间数列的速度指标	181
一、发展速度	181
二、增长速度	182
三、平均发展速度	185
四、平均增长速度	188
第四节 时间数列的分析	188
一、时间数列结构分析的意义	188
二、时间数列的构成因素	188
三、时间数列构成因素的结合方式	190
四、长期趋势的测定方法	190
五、季节变动趋势分析	203
六、循环变动分析	206
七、不规则变动的测定方法	207
复习思考题	208
第九章 统计指数	212
第一节 统计指数概述	212
一、统计指数的概念	212
二、指数的种类	212
三、指数的性质	214
四、指数的作用	214
五、指数编制的基本问题	215
第二节 综合指数	217
一、综合指数的定义及特点	217
二、综合指数的编制原理	217
三、综合指数的具体形式	218
第三节 平均指数	224
一、平均指数的概念及特点	224
二、加权平均数指数的编制原理	224
三、加权平均数指数的具体形式	225
四、综合指数与平均指数的联系和区别	228
第四节 平均指标指数	229
一、平均指标指数的概念	229
二、平均指标指数的编制步骤	229
第五节 指数体系与因素分析	230
一、指数体系的含义及其作用	230

二、因素分析	231
三、总量指标因素分析	232
四、平均指标变动的因素分析	236
第六节 几种常用的经济指数	238
一、工业生产指数	239
二、零售价格指数	240
三、消费价格指数	242
四、股票价格指数	243
五、农副产品收购价格指数	244
六、产品成本指数	244
复习思考题	245

第十章 统计分析..... 248

第一节 统计分析概述	248
一、统计分析的概念	248
二、统计分析的意义	248
三、统计分析的种类	249
第二节 统计分析方法和步骤	250
一、统计分析方法	250
二、统计分析的步骤	255
第三节 统计分析报告	256
一、统计分析报告的概念	256
二、统计分析报告的特点	256
三、统计分析报告的写作原则	258
四、统计分析报告的结构	258
复习思考题	259

第十一章 国民经济统计概述..... 260

第一节 国民经济统计的基本内容	260
一、国民经济统计的意义	260
二、我国国民经济统计的内容体系	261
三、基本核算表的主要内容	262
第二节 国民经济统计的主要分类	263
一、国民经济统计分类的一般问题	263
二、国民经济的一般理论分类	264
三、国民经济行业分类与机构部门分类	266
四、常住单位与非常住单位的区分	270
第三节 国民经济统计的常用指标	271
一、国民经济统计指标体系	271
二、国民经济生产指标	271

三、收入分配和收入使用指标	274
四、资产负债和国民财富指标	275
复习思考题	277
第十二章 常用统计软件及应用	279
第一节 统计软件的类型	279
一、统计软件的选择	279
二、常用统计软件的类型	279
第二节 SPSS 软件的简单应用	281
一、SPSS 软件的特点	281
二、利用 SPSS 进行统计处理的基本过程	282
三、SPSS 数据文件的建立	283
第三节 SAS 系统的简单应用	287
一、SAS 系统的特点	287
二、SAS 软件的操作	288
复习思考题	292
附录：常用统计表	293
参考文献	325

第一章 绪 论

第一节 统计学的产生与发展

统计学的产生和发展是伴随着资本主义古典哲学、古典政治经济学和空想社会主义的产生而发展起来的。一般公认,统计学从17世纪60年代算起,到现在已有近400年的历史。史料说明,统计学起源于定量地说明和研究社会经济问题。它是帮助决策、辅助管理的工具。统计学的发展过程大体可划分为以下四个阶段。

第一个阶段称之为“城邦政情”阶段。

“城邦政情”(Matters of State)阶段始于古希腊的亚里斯多德撰写的“城邦政情”或“城邦纪要”。他一共撰写了150余种纪要,其内容包括各城邦的历史、行政、科学、艺术、人口、资源和财富等社会及经济情况的比较与分析,具有社会科学特征。“城邦政情”式的统计研究延续了一两千年,直至17世纪中叶才逐渐被“政治算术”这个名词所替代,并且很快被演化为“统计学”(Statistics)。统计学依然保留了城邦(State)这个词根。

第二个阶段称之为“政治算术”阶段。

“政治算术”(Political Arithmetic)阶段与“城邦政情”阶段并没有很明显的分界点,两者也没有根本的差别。“政治算术”的特点是统计方法与数学计算和推理方法相结合,分析社会经济问题的方式更加注重运用定量分析方法。可以用1690年英国古典政治经济学的奠基人威廉·佩蒂(William Petty, 1623~1687)出版《政治算术》一书作为这个阶段的起始标志。威廉·佩蒂在书中强调了自己的立论方法,他说:“我进行这种工作所使用的方法,在目前还不是常见的。因为我不采用比较级或最高级的词语进行思辨式的议论,相反地采用了这样的方法(作为我很久以来就想建立的政治算术的一个范例),即用数字、重量和尺度来表达自己想说的内容,只进行诉诸人们感觉的议论,借以考察在自然中有可见根据的原因。”

用数字、重量和尺度量化社会经济现象的方法是近代统计学的重要特征。因此,威廉·佩蒂的《政治算术》被后来的学者评价为近代统计学的来源,威廉·佩蒂本人也被评价为“近代统计学之父”。刺激近代统计学发展的主要原因是战争和工业革命。当时,政治家需要有人专门致力于收集和分析社会、经济及政治方面的统计资料,帮助他们决策与管理国家。威廉·佩蒂写作《政治算术》大约在1671~1676年之间,正值第三次英荷战争(1672~1674年)期间。与此同时,他还撰写了《爱尔兰的政治解剖》一书。在威廉·佩蒂的心目中,政治算术和政治解剖这两种方法是二位一体的。有些学者认为《政治算术》是“比较统计”(Comparatives Statistic),《爱尔兰的政治解剖》是“个别统计”(Einzel Statistic)。

构成“政治算术”方法的核心是“数字”。佩蒂在书中使用的数字有三类:第一类是对社会经济现象通过统计调查和经验观察得到的数字,由于受到历史条件的限制,书中通过严格的统计调查得到的数据少,根据经验得出的数字多;第二类是运用某种数学方法推算出来的数字。佩蒂在书中多次使用这类数据,其推算方法可分如下三种:

- (1) 以已知数或已知量为基础,循着某种具体关系进行推算的方法;
- (2) 通过运用数字的理论性推理来进行推算的方法;

(3) 以平均数为基础进行推算的方法。

第三类是为了进行理论性推理而采用的示例性数字。佩蒂把这种运用数字和符号进行的推理称之为“代数的算法”。从佩蒂使用数据的方法看,“政治算术”阶段的统计学已经比较明显地体现了“收集和分析数据的科学与艺术”的特点,而且统计实证方法和理论分析方法浑然一体,这种方法即使是现代统计学也依然继承。

另外,英国的约翰·格朗特(John Graunt, 1620~1674)在1662年出版了《关于死亡表的自然观察和政治观察》。他通过大量观察,研究并发现了人口与社会现象中重要的数量规律性。法国的帕斯卡尔(B. Pascal, 1623~1662)和费马特(Pierre de Fermat, 1601~1665)在数学家们对机会游戏研究的基础上,通过通信的方式,将赌博中出现的各种具体问题归结为一般的概率原理,为概率论和统计学的发展奠定了重要的基础。特别值得提出的是比利时统计学家凯特勒(Quetelet. L. A. J, 1796~1874),他对统计学理论最大的影响是把概率与统计学相结合,从而提出了关于统计学的新概念,他的主要功绩在于使统计方法获得普遍应用,他对各种科学均有造诣,像天文学、数学、物理学、人类学、行政、生命和社会统计学、气象学等,并努力使人们把统计学理解为一种可用于任何学科的一般研究方法。在“政治算术”阶段出现的统计学与数学的结合趋势逐渐发展形成了“统计分析科学”。

第三个阶段称之为“统计分析科学”阶段。

19世纪末,欧洲大学里开设的“国情纪要”或“政治算术”等课程名称逐渐消失,取而代之的是“统计分析科学”(Science of Statistical Analysis)课程。当时的“统计分析科学”课程的内容仍然是分析研究社会经济问题。“统计分析科学”课程的出现是现代统计学发展阶段的开端。

经过历代统计学家的努力和发展,到19世纪末建成了古典统计学的基本框架。1908年,英国的戈塞特(William Seely Gosset, 1876~1937)用笔名 Student 发表了关于 t 分布的论文,这是一篇在统计学发展史上划时代的文章。它创立了小样本代替大样本的方法,开创了统计学的新纪元。现代统计学的代表人物首推比利时统计学家奎特莱(Adolphe Quelet),他将“统计分析科学”广泛应用于社会科学、自然科学和工程技术科学领域,因为他深信统计学是可以用于研究任何科学的一般研究方法。

后来,统计学家费希尔(Ronald Aylmer Fisher, 1890~1962)给出了 F 统计量、极大似然估计、方差分析等;内曼(J. Neyman)和皮尔逊(Egon S. Pearson)提出了置信区间估计和假设检验;沃尔德(A. Wald)发明了序贯抽样和统计决策函数,到20世纪50年代,构筑了现代统计学的基本框架。20世纪50年代以后,统计学进入了一个全新的发展阶段,新的研究领域层出不穷,像多元统计分析、非参数统计、现代时间序列分析和线性统计模型等。

第四个阶段称之为“统计学帝国主义”阶段(Imperialistic Stage of Statistics)。

从世界范围看,自20世纪60年代以后,统计学的发展有三个明显的趋势:第一,随着数学的发展,统计学依赖和吸收的数学方法越来越多;第二,向其他学科领域渗透,或者说以统计学为基础的边缘学科不断形成;第三,随着统计学的应用日益广泛和深入,以及受计算机和新兴科学的影响,统计学越来越依赖于计算技术,成为数量分析的方法论科学。这一时期统计学的研究和应用范围越来越广。例如,海洋学中水面的高程、电磁能(特别是无线电波)的流量、工业过程的状态和生物的状态都能用计算机收集,没有计算机这些是不实际或不可能的。计算机节省了大量人力,特别是在同样输出的重复计算上,比如在计算多元分析的相关系数和其他检验统计量时。由于计算机软件 SPSS 和 SAS 可用于所有的通常检验,

特别是关于方差分析,则节省更多的时间。由于利用软件很方便,2003年度诺贝尔经济学奖授予了两位著名计量经济学家罗伯特·恩格尔(Robert F.Engle)和克莱夫·格兰杰(Clive Granger)。20世纪80年代,两位获奖者发明了新的统计方法来处理许多经济时间数列中两个关键属性:时变性(Time-Varyingvolatility)和非平稳性(Nonstationarity)。恩格尔的研究方向主要是利率、汇率和期权的金融计量分析,提出谱分析回归等创新性统计方法;格兰杰的研究涉及统计和经济计量学,特别是时间序列分析、预测、金融、人口统计学和方法论等领域。这足以证明统计方法应用的领域越来越广泛,可以应用于各行各业的数据分析,使统计学成了一门“万能”的方法论学科。美国《SCIENCE》杂志有一篇文章列出近百年来最有用的科学,统计学位居前10名。

许多新的分支或专门化的应用已经得到了发展,如决策论、时间序列、多元分析、经济计量学、博弈论、临床试验、非参数推断、序贯分析、数学生物分类学及可靠性分析,包括科克伦(W. G. Cochran, 1909~1980)的实验设计理论、安德森(Th. W. Anderson)的复变数分析等等。数理统计及其应用正在继续发展和扩大。

萨尔斯伯格(David Salsburg)在《统计,改变了世界》的序言中讲到:17世纪以来,统计学早已跨越政府统计的范畴,静悄悄地渗入数理、生物、人文和工程等领域,甚至在日常生活中也处处可见统计的踪迹,像民意调查和企业成效的评估等,无一不与统计相关。统计学,事实上已经改变了我们对自然、心智以及社会的看法,这场静默革命深刻且广泛,可惜的是一般人虽然身在其中,却认识不深。

第二节 统计学的性质和特点

一、统计与统计学的涵义

对于“统计”一词,我们其实很难给出一个简单的定义,一般可以有三种不同的涵义,即统计工作、统计资料和统计学。

1. 统计工作

统计工作是统计设计、统计调查、统计整理、统计分析和研究统计数据资料的工作过程。统计工作在人类历史上出现得比较早。随着历史的发展,统计工作逐渐发展和完善起来,并成为国家、企业和个人及科研单位认识与改造客观世界和主观世界的一种有力工具。

一般而言,任何统计工作都要经过统计设计、统计调查、统计整理和统计分析四个阶段。

第一个阶段是统计设计。它是统计工作在实际展开前所做的计划性安排,既包括统计方案的全面设计,又包括人、财、物等各种资源的安排。一项统计工作能否顺利进行与统计设计是否全面、周密有很大的关系。

第二个阶段是统计调查。此阶段充分反映了统计工作的特点,是深入实际调查研究的重要环节。统计调查与其他调查活动的不同点在于:精心设计的统计调查方法更能保障调查活动起到事半功倍的效果。统计调查的方法分为普查、统计报表制度、重点调查、典型调查和抽样调查五种,这五种方法各有特点。普查活动搜集的资料最全面、详细、准确,但缺点也很明显,那就是费钱、费时且费人力,因为普查要求对每一个调查对象都要调查到,比如人口普查,国家每次人口普查都需要消耗大量的社会资源。重点调查和典型调查是我们短期迅

速估查的好方法,但结果难以定量化,而且容易产生无法控制的误差,随意性很大。抽样调查是最经常采用的一种调查方法,原因在于它具有时效性强、低成本、高质量和科学性的特点,虽然抽样调查的结果存在误差,但这个误差是我们可以接受和控制的。抽样调查有一整套科学方法,是统计学的重要分支。当然,要切实做好统计调查这项工作也决非易事,除了要掌握方法外,调查人员的责任心至关重要。

第三个阶段是统计整理。对通过统计调查搜集到的原始数据、统计资料,根据研究的需要进行分类、汇总等。统计整理是一个对统计调查结果去粗取精、去伪存真的过程,就是要对统计调查搜集到的原始数据进行“把关”,修匀和剔除一些无效问卷。统计整理后的结果就是我们通常看到的各种统计资料了。

第四个阶段是统计分析。通过分析搜集到的统计资料来揭示其内在的一些规律性的特点。这一阶段充分体现了统计学的精华,主要表现为对各种统计分析方法的应用,如统计指标、时间序列分析、统计指数和回归分析等。

我国县以上各级政府均设有专门的统计部门和机构,自下而上成体系的机构主要是国家统计局系统,包括县以上各级统计局和三支统计调查的专业队伍(城市社会经济调查队、农村社会经济调查队和企业社会经济调查队),这是我国社会经济发展的数据资料搜集、整理和分析的专业化国家队。他们是每年发布大量的社会经济发展的宏观数据的权威部门。

2. 统计资料

统计资料(Statistical Data)是指可用以推导出某项结论的一些事实或数字。统计资料一般由元素、变量和观测值三部分组成。

(1) 元素(Element)。我们所研究的对象由各元素组成,统计资料就是关于各元素特征的信息。

(2) 变量(Variable)。变量是关于元素的一种属性或特征。

如表 1-1 所示,年龄是各员工的一个特征,这些特征对于不同元素取不同的结果,所以称为变量。其中结果用数字表示的,称为定量变量(Quantitative Variable);不可用数字表示的,称为定性变量(Qualitative Variable)。

表 1-1 员工个人资料

姓 名	性 别	年 龄	身高 (m)	体重 (kg)	民 族	公司服务年限	受教育年限
甲	男	33	1.85	65	汉	3	18
乙	女	25	1.65	55	回	2	16
丙	男	26	1.72	60	满	1	15
丁	女	35	1.60	53	回	4	16
戊	男	32	1.83	68	汉	2	19

在表 1-1 中,年龄、身高、体重、公司服务年限和受教育年限是定量变量;性别、民族为定性变量。

(3) 观测值(Observation)。一次观测的结果是指资料中关于某一元素所有变量的信息。表 1-1 中的每行都表示一次观测的结果。

当一项统计研究工作开始时,首先要明确所研究的问题以及为此需要哪些资料,随之就要考虑如何收集以及到哪里收集这些资料,是间接引用现成的资料呢,还是直接自己动手。

若有现成资料可以引用,当然应尽量引用。这样不仅可以节省费用和时间,而且因为某

些资料（如地域广、历史性或专业性很强的）自己直接收集通常是不可能的。因此，要求统计工作者必须了解现成的统计资料的情况，如各级政府公报、年鉴、期刊杂志和各种专业数据库等，以便及时引用。同时还应掌握现代的检索方法。

当所需要的资料没有适当的、现成的可以引用时，就需要自己动手直接收集。为了获得所需要的资料，在调查之前先要明确采取哪种研究方式。一般情况下，有两种取得统计资料的研究方式：实验式和非实验式（或称观察式）。

实验式收集资料的方式就是设计一种统计实验，根据某些可以控制的因素的变化得到关于这些因素对研究的变量的影响信息，而将那些控制因素以外的条件保持不变，或将控制因素以外的其他因素的影响用随机化的方法加以平衡抵消。例如，泰勒（F. W. Taylor）的科学管理理论中的工作定额原理就是用实验式方法获得工人合理的日工作量的资料。

非实验式（或称观察式的）收集资料的方式是对影响所研究的变量的任何因素都不加限制或控制，直接观察变量的取值。例如，在收集气象资料时，一般是用非实验式的方法。在研究商品价格的变化对商品需求量的影响时，显然任意改变价格的方法是不能考虑的。政府在研究国民收入对政府财政收入的影响时，也不能考虑采取实验式方法，这类问题只能从现有现象中去收集资料，然后分析人员根据分析目的选出有用的结果。

3. 统计学

“统计”一词，英语为 Statistics，用作复数名词时，意思是统计资料，作单数名词时，指的是统计学。统计一词最早出现于拉丁语的 Status，意思指各种现象的状态和状况。由这一语根组成意大利语 Stato，表示“国家”的概念，也含有国家结构和国情知识的意思。根据这一语根，最早作为学名使用的“统计”是在 18 世纪德国政治学教授亨瓦尔（G. A. Hewall）在 1749 年所著《近代欧洲各国国家学纲要》一书的序言中，他把国家学名定为“Statistika”（统计）这个词，原意是指“国家显著事项的比较和记述”或“国势学”，认为统计是关于国家应注意事项的学问。此后，各国相继沿用“统计”这个词，并把这个词译成各国的文字，法国译为 Statistique，意大利译为 Statistica，英国译为 Statistics，日本最初译为“政表”、“政算”、“国势”和“形势”等，直到 1880 年在太政官中设立了统计院，才确定以“统计”二字正名。1903 年（清光绪二十九年），由钮永建和林卓南等人翻译了横山雅南所著的《统计讲义录》一书，把“统计”这个词从日本引入我国。1907 年（清光绪三十三年），彭祖植编写的《统计学》在日本出版，同时在国内发行，这是我国最早的一本统计学书籍。“统计”一词就成了记述国家和社会状况的数量关系的总称。事实上，统计学发展到现代，其涵义已经发生了很大的变化。

那么统计学的主要内容究竟是什么呢？大多数的科学家都有过分析数据的经验，也很可能会遇到下列的问题：如何搜集到有价值的数据？如何组织和解释所搜集的资料？如何分析并得出适当的推论？以及推论的可信度有多高？这些都是统计学研究的主要范畴。这其中当然包括资料从搜集、分析到推论的整个过程中所需具备的知识。诸如民意调查、医学的临床试验及农场质量的改良试验等，都是运用统计学理论的常见例子。

一般来说，统计学是对研究对象的数据资料进行搜集、整理、分析和研究，以显示其总体的特征和规律性的学科。统计学的研究对象是客观事物的数量特征和数据资料。统计学是以搜集、整理、分析和研究等统计技术为手段，对所研究对象的总体数量关系和数据资料去伪存真、去粗取精，从而达到显示、描述和推断被研究对象的特征、趋势及规律性的目的。简言之，统计学是以少量的数据（称为样本）所提供的信息来推断欲研究对象（称为母体）