

翟永平/著



统计分析及 相关软件应用

T

*Tongji Fenli Ji
Xiang Guan Ruan Jian Ying Yong*



经济科学出版社
Economic Science Press

统计分析及相关软件应用

翟永平 著

经济科学出版社

责任编辑：王长廷 刘 莎

责任校对：王肖楠

版式设计：代小卫

技术编辑：邱 天

图书在版编目 (CIP) 数据

统计分析及相关软件应用/翟永平著. —北京：经济科学出版社，2010.3

ISBN 978 - 7 - 5058 - 9062 - 6

I. 统… II. 翟… III. 统计分析 - 应用软件
IV. C819

中国版本图书馆 CIP 数据核字 (2010) 第 019387 号

统计分析及相关软件应用

翟永平 著

经济科学出版社出版、发行 新华书店经销

社址：北京市海淀区阜成路甲 28 号 邮编：100142

总编部电话：88191217 发行部电话：88191540

网址：www. esp. com. cn

电子邮件：esp@ esp. com. cn

北京密兴印刷厂印装

880 × 1230 32 开 10.25 印张 270000 字

2010 年 3 月第 1 版 2010 年 3 月第 1 次印刷

ISBN 978 - 7 - 5058 - 9062 - 6 定价：30.00 元

(图书出现印装问题，本社负责调换)

(版权所有 翻印必究)

目 录

第一章 统计学概述	1
第一节 什么是统计学	1
第二节 统计学的基本问题	14
第三节 统计数据处理途径	24
第二章 统计软件介绍	30
第一节 SPSS 软件分析功能	30
第二节 Excel 统计分析功能	33
第三章 统计资料的收集	39
第一节 数据类型	40
第二节 统计资料的收集	42
第三节 问卷设计	50
第四章 统计资料的初步分析	57
第一节 资料整理	57
第二节 统计表与统计图	66
第三节 Excel 在数据初步分析中的应用	76
第五章 数据的描述性分析	85
第一节 集中趋势的描述	85

第二节	离散程度的描述	100
第三节	分布偏态与峰度的测度	107
第四节	Excel 在描述分析中的应用	110
第五节	SPSS 在描述分析中的应用	115
第六章	抽样与参数估计	122
第一节	抽样推断的基本概念与原理	122
第二节	参数估计	129
第三节	抽样组织方式及其参数估计	139
第四节	样本容量的确定	144
第五节	Excel 在抽样与参数估计中的应用	147
第七章	假设检验	152
第一节	假设检验概述	152
第二节	一个总体参数的检验	158
第三节	两个总体参数的检验	162
第四节	总体方差的检验	166
第五节	Excel 在假设检验中的应用	168
第八章	方差分析	174
第一节	方差分析概述	174
第二节	单因素方差分析	177
第三节	双因素方差分析	183
第四节	Excel 在方差分析中的应用	190
第九章	相关与回归分析	199
第一节	相关分析	199
第二节	一元线性回归分析	205
第三节	多元线性回归分析	216

第四节 非线性模型·····	221
第五节 Excel 在回归分析中的应用 ·····	223
第十章 时间序列分析·····	230
第一节 时间序列概述·····	231
第二节 时间序列的描述性分析·····	234
第三节 因素分析·····	242
第四节 Excel 在时间序列分析中的应用 ·····	263
第十一章 统计指数·····	270
第一节 指数概念·····	270
第二节 总指数的编制方法·····	273
第三节 指数体系与因素分析·····	281
第四节 几种常用的经济指数·····	288
第五节 Excel 在指数分析中的应用 ·····	296
附录·····	306
主要参考文献·····	318

第一章 统计学概述

现代社会中，我们每天都会面对大量的数据，看看电视，翻翻报纸，数据无处不在。要使这些数据变为对你有用的信息，就要对这些数据作一些处理和分析。怎样处理和分析数据，这正是统计学的用武之地。

第一节 什么是统计学

一、统计与统计学的含义

日常生活中，我们经常会接触到“统计”这个词，在人们的一般认识中，“统计”就是“计数”。其实，最早的统计就是一种计数活动，如大家熟知的“结绳记事”。小至一个人、家庭，大至一个企业、国家都有计数的任务，一个月的收入、一年的利润都会是我们经常关心盘算的。在外语中，“统计”一词与“国家”一词来自同一词源。因此，可以说自从有了国家就有了统计实践活动。最初，统计只是为统治者管理国家的需要而搜集资料，弄清国家的人力、物力和财力，作为国家管理的依据。

今天，“统计”一词已被人们赋予多种含义，因此很难给出一个简单的定义。在不同场合，统计一词可以具有不同的含义。它可以指统计数据的搜集活动，即统计工作；也可以指统计活动的结

果，即统计数据资料；还可以是指分析统计数据的方法和技术，即统计学。

（一）统计工作

统计工作是为了管理或认识的需要，对社会经济现象和自然现象进行数量收集的活动。统计工作从人类产生、劳动开始，也就开始了，“结绳记事”就是最早的统计工作。统计工作的延续与发展使得其范围越来越大、经验积累得越来越多，客观上为统计学的产生和发展奠定了基础。统计工作，可以简称为统计。例如，某统计师在回答自己的工种时，会说我是干统计的。这里所说的统计指的就是统计工作。

（二）统计资料

统计资料是统计工作过程中所取得的各项数字资料以及与之相关信息总称。统计资料是统计工作取得的、能够说明所研究对象的数据。不管是个人、集体和社会，还是国家、部门和事业、企业、公司及科研机构，都离不开统计数据资料。个人要进行学习、工作和家政管理，需要对有关的统计数据资料进行搜集和分析，以指导自己的学习、工作和生活；公司和企业要管理好生产和销售，必须进行市场调研、生产控制、质量管理、成本评估等，这就需要对有关的生产资料、市场资料、成本资料、质量数据等进行搜集、整理、分析和研究；国家要进行经济建设和社会发展，更离不开有关国民经济和社会发展的统计资料。

图 1-1 所示的中国统计出版社出版的每年一册的《中国统计年鉴》和国家统计局每年初公布的《国民经济与社会发展统计公报》等都是统计数据资料，也可称为统计。

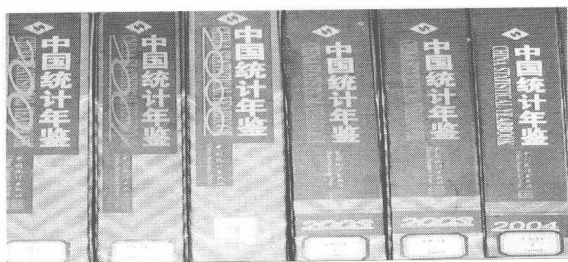


图 1-1 中国统计资料的获取渠道

(三) 统计学

目前,随着统计方法在各个领域的应用,统计学已发展成为具有多个分支学科的大家族。因此,要给统计学下一个普遍接受的定义是十分困难的。在本书中,我们对统计学做如下解释:统计学是一门收集、整理和分析统计数据的方法科学,其目的是探索数据的内在数量规律性,以达到对客观事物的科学认识。统计学的定义突出了统计学研究对象的三个方面:第一,收集数据;第二,整理数据;第三,分析数据。

统计数据的收集是取得统计数据的过程,它是进行统计分析的基础。离开了统计数据,统计方法就失去了用武之地。如何取得所需的统计数据是统计学研究的内容之一。

统计数据的整理是对统计数据的加工处理过程,目的是使统计数据系统化、条理化,符合统计分析的需要。数据整理是介于数据收集与数据分析之间的一个必要环节。

统计数据的分析是统计学的核心内容,它是通过统计描述和统计推断的方法探索数据内在规律的过程。

【特别提示】在英文中,“Statistics”一词有两个含义:当它以单数名词出现时,表示作为一门科学的统计学;当它以复数名词出现时,表示统计数据或统计资料。

可以看出,统计学与统计数据之间有着密不可分的关系。统计

学是由一套收集和处理统计数据的方法所组成的，这些方法来源于对统计数据的研究，目的也在于对统计数据的研究。统计数据不用统计方法去分析也仅仅是一堆数据而已，无法得出任何有益的结论。

统计数据不是指单个的数字，而是由多个数据构成的数据集。单个的数据显然用不着统计方法进行分析，仅凭一个数据点，我们也不可能得出事物的规律，只有经过对同一事物进行多次观察或计量得到大量数据，才能利用统计方法探索出内在的规律性。

【例 1.1】从电讯局或网络公司的业务记录中获得了相关数据，加以汇总后见表 1-1，反映每天某个时间点上，在线用户占注册用户比例。

表 1-1

用户通常上网时间

时 间	上网比例 (%)	时 间	上网比例 (%)	时 间	上网比例 (%)
0 点	14.5	8 点	11.8	16 点	31.2
1 点	5.6	9 点	22.5	17 点	27.7
2 点	4.2	10 点	24.6	18 点	22.1
3 点	2.6	11 点	22.6	19 点	31.3
4 点	2.5	12 点	22.6	20 点	44.0
5 点	1.6	13 点	24.3	21 点	43.9
6 点	1.7	14 点	28.6	22 点	35.0
7 点	2.7	15 点	31.4	23 点	22.2

资料来源：中国互联网络信息中心，《中国互联网络发展状况统计报告（2003）》。

统计表固然能清晰地反映数据，但更直观的是统计图，根据上面的数据我们用 Excel 图表功能很快就能绘制出图形来，见图 1-2。

从图表中我们清楚地看到这样一个统计规律：晚上的八九点钟是上网的高峰时间，而清晨五六点钟上网的人最少。

统计学研究数据整理与信息展现，更注重对所获得资料的分析。用各种各样的方式对数据进行处理，让貌似静止的数据资料能开口说话，让它为我们的决策提供依据，给我们带来经济效益。

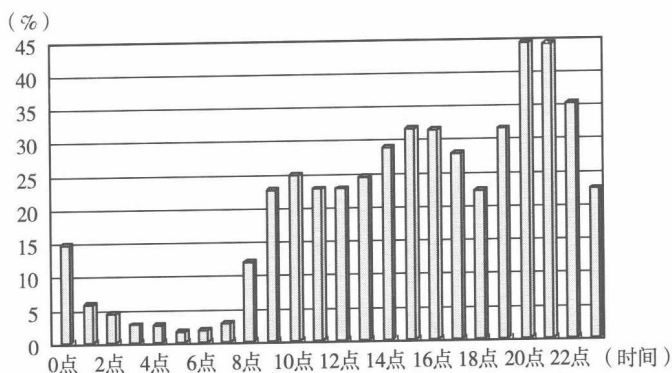


图 1-2 互联网上各时间段的使用比例

二、统计学的发展历程

人类的统计实践是随着计数活动而产生的。因此，对统计发展的万史可追溯到远古的原始社会。在我国，公元前 2250 年大禹治水时，根据地理位置、人口及物产和贡赋的多少，将全国分为九个州，汇编成《禹贡九州篇》，从而形成统计的雏形，为统计学的萌芽奠定了基础。但是，将人类的统计实践上升到理论予以总结和概括成为一门系统的科学——统计学，却是近代的事情，距今只有三百多年的历史。回顾一下统计科学的渊源及其发展过程，对于我们了解统计学的研究对象和性质，学习统计学的理论和方法，提高我们的统计实践和理论水平，都是十分必要的。

从统计学的产生和发展过程来看，大致可以划分为三个时期：统计学的萌芽期、统计学的近代期和统计学的现代期。

（一）统计学的萌芽期（17 世纪中叶至 18 世纪）

1. 国势学派

国势学派产生于 17 世纪的德国，代表人物是康令（H. Conring）、阿亨瓦尔（G. Achenwall）。康令是德国一所大学的教授，他们在大

学中开设了一门新课程，最初叫做“国势学”。他们所做的工作主要是对国家重要事项的记录，因此又被称为记述学派。这些记录记载着关于国家、人口、军队、领土以及资源财产等事项，偏重于事件的叙述，主要用的是文字叙述的形式，基本上没有量的描述与分析。

国势学派对统计学的最大贡献就是它提出了一个世界公认的名词“统计学”（Statistics）。另外该学派在研究各国的重要事件时，采用了系统对比的方法来反映各国实力的强弱，统计学分析方法中的“对比”的思想实源于此。有人评价该学派是“有统计之名，无统计之实”的学派。

2. 政治算术学派

该学派起源于17世纪的英国，同是研究各国的国情、国力，英国的学者们却采用了数量分析的方法，因此，从严格意义上来说，政治算术学派作为统计学的开端更为合适。代表人物是威廉·配第（William Petty）和约翰·格朗特（J. Graunt）。配第在其著作《政治算术》中，对当时的英国、荷兰、法国之间的财富进行数量上的计算和比较，开用数量方法研究社会经济现象之先河。在这个意义上，马克思称配第是“政治经济学之父，在某种程度上也可以说是统计学的创始人”。

格朗特通过对伦敦市五十多年的人口出生和死亡资料的计算，写出了第一本关于人口统计的著作《对死亡率公报的自然观察和政治观察》，统计的含义也因此从记述转变为专指用数量来说明国家的重要事件。

政治算术学派在统计发展的历史上有着重要的地位，它以数量分析为特征，研究客观现象的数量方面，就这点来说，政治算术应该是统计学的起源。这个学派在搜集资料方面，较明确地提出了大量观察法、典型调查、定期调查等思想；在处理资料方面，较为广泛地运用了分类、制表及各种指标来浓缩与显现数量资料的内涵信息。同时，政治算术学派第一次运用可度量的方法，力求把自己的

论证建立在具体的、有说服力的数字上面，依靠数字来解释与说明社会经济生活。然而，政治算术学派毕竟还处于统计发展的初创阶段，它只是用简单的、粗略的算术方法对社会经济现象进行计量和比较。有人评价该学派是“有统计之实，无统计之名”的学派。

（二）统计学的近代期（18 世纪末至 19 世纪末）

1. 数理统计学派

最初的统计方法是随着社会政治和经济的需要而初步得到发展的。直到概率论被引进之后，才逐渐形成一门成熟的科学。在统计发展史上，最初把古典概率论引进统计学领域的是法国天文学家、数学家、统计学家拉普拉斯（P. S. Laplace）。他发展了对概率论的研究，阐明了统计学的大数法则，并进行了大样本推断的尝试。

随着资本主义经济的发展，统计被应用于社会经济的各个方面，统计学逐步走向昌盛。比利时统计学家、数学家、天文学家凯特勒（Adolphe Quetelet）完成了统计学和概率论的结合。从此，统计学开始进入更为丰富发展的新阶段。国际统计学界有人称凯特勒为“统计学之父”，就在于他发现了大量现象的统计规律性和开创性地应用了许多统计方法。1841 年他出任比利时中央统计委员会会长，1851 年积极筹备国际统计学会组织，并任第一届国际统计会议主席；他在著作《社会物理学》中利用大数法则论证了社会生活中的随机偶然现象贯穿着必然的规律性，他运用概率论原理提出了著名的“平均人”的概念，计算人类自身各性质标志的平均值，通过“平均人”来探索社会规律，他认为，社会所有的人与“平均人”的差距越小，则社会矛盾就越缓和。凯特勒被统计学界称为“国际统计会议之父”和“近代统计学之父”，其贡献就是发现了大量现象的统计规律性和开创性地应用了许多统计方法，为数理统计学的发展奠定了基础。

同时，凯特勒也是数理统计学派的奠基人，因为数理统计就是

在概率论的基础上发展起来的。随着统计学的发展,对概率论的运用逐步增加;同时,自然科学的迅速发展和技术的不断进步,对数理统计方法又提出了进一步的要求。这样,数理统计学就从统计学中分离出来自成一派。由于这一学派主要是在英美等国发展起来的,故又称英美数理统计学派。

2. 社会统计学派

社会统计学派由德国大学教授克尼斯(K. G. A. Knies)首创,主要代表人物为恩格尔(C. L. E. Engel)和梅尔(G. V. Mayr)。他们认为,统计学的研究对象是社会现象,目的在于明确社会现象内部的联系和相互关系;在研究过程中,要用全面调查,也可以适量地使用抽样调查。恩格尔在1895年发表的《比利时工人家庭的生活费》一文中,提出了著名的“恩格尔法则”,从中引申的“恩格尔系数”,作为衡量生活水平的标准,至今仍被沿用。

各个学派的传承与发展,使今天的统计学形成这样的格局:一是以社会经济问题为主要研究对象的社会经济统计;二是以方法 and 应用研究为主的数理统计。从学科的角度看,前者从属于应用经济学;后者从属于数学。

(三) 统计学的现代期

统计学的现代期是自20世纪初到现在的数理统计时期。20世纪20年代以来,数理统计学发展的主流从描述统计学转向推断统计学。19世纪末和20世纪初的统计学主要是关于描述统计学中的一些基本概念、资料的搜集、整理、图示和分析等,后来逐步增加概率论和推断统计的内容。直到20世纪30年代,费希尔的推断统计学才促使数理统计进入现代范畴。

现在,数理统计学的丰富程度完全可以独立成为一门学科,但它也不可能完全代替一般统计方法论。传统的统计方法虽然比较简单,但在实际统计工作中运用仍然极广,正如四则运算与高等数学的关系一样。不仅如此,数理统计学主要涉及资料的分析和推断方

面,而统计学还包括各种统计调查、统计工作制度和核算体系的方法理论、统计学与各专业相结合的一般方法理论等。由于统计学比数理统计在内容上更为广泛,因此,数理统计学相对于统计学来说不是一门并列的学科,而是统计学的重要组成部分。

从世界范围看,自 20 世纪 60 年代以后,统计学的发展有几个明显的趋势: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,随着统计学应用日益广泛和深入,特别是借助电子计算机后,统计学所发挥的功效日益增强。首先统计学从面对小批量的数据转变为面对海量数据,因此使用计算机统计分析软件对数据进行处理成为必然;在某些领域,甚至约定俗成必须使用著名统计分析软件 SAS,否则无法认可分析结果的准确性。第四,统计学的作用与功能已从描述事物现状、反映事物规律,向抽样推断、预测未来变化方向发展。它已从一门实质性的社会性学科,发展成为方法论的综合性学科。

【特别提示】2003 年诺贝尔经济学奖授予了著名计量经济学家恩格尔 (Robert F. Engle) 和格兰杰 (Clive Granger),两位首创了新的统计方法来处理许多经济时间数列中时变性和非平稳性,涉及金融、人口等,这证明了统计方法应用的领域越来越广泛。统计学可以应用于各行各业的数据分析,这使得它成为一门“万能”的方法论学科。

三、统计学的研究方法

1. 大量观察法

大量观察法是统计学所特有的方法。所谓大量观察法,是指对所研究的事物的全部或足够数量进行观察的方法。统计学的研究对象是客观现象总体的数量方面,这个总体是由同一性质的许多个体单位构成的,由于客观现象的错综复杂性,这些个体的特征和数量

表现会存在很大的差异,但是如果我们综合大量的个体,则偶然性的数量差异又会相互抵消,而清晰地呈现出现象的数量特征。因此,统计研究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。

大量观察法的数学依据是大数定律。大数定律是随机现象的基本规律。大数定律的一般概念是:在观察过程中,每次取得的结果不同,这是由偶然性所致的,但大量、重复观察结果的平均值却几乎接近确定的数值。狭义的大数定律就是指概率论中反映上述规律性的一些定理,表述平均数的规律性与随机现象的概率关系。

大数定律的本质意义在于经过大量观察,把个别的、偶然的差异性相互抵消,而必然的、集体的规律性便显示出来。例如,当我们观察个别家庭或少数家庭的婴儿出生时,生男生女的比例极为参差不齐,有的是生男不生女,有的是生女不生男,有的是女多男少,有的是男多女少,然而经过大量观察,男婴、女婴的出生数则趋向均衡。也就是说,观察的次数愈多,离差的差距就愈小,或者说频率出现了稳定性。这就表明,同质的大量现象是有规律的,尽管个别现象受偶然性因素的影响出现偏差,但观察数量达到一定程度就呈现出规律性,这就是大数定律的作用。

2. 统计描述法

统计描述法是对由调查或实验得到的统计数据资料进行整理、归类,计算出各种能反映总体数量特征的综合指标,并加以分析研究,从而得出需要的数据资料信息,用表格、图形和统计指标数值来表示的统计方法。统计描述是统计研究的基础,它为统计推断、统计咨询和统计决策提供必要的统计数据资料。统计描述法的具体方法是应用统计分组法、综合归纳法等得到现象总体的数量特征,反映客观事物的内在数量规律性,以达到进行统计分析和研究的目的。

统计分组法是将总体中的个体分为若干个组,以研究总体内部差异的一种常用统计方法。通过统计分组法,可以研究总体中不同

类型的性质以及构成、分布特征。例如，三次产业的划分，可以分析研究三次产业的结构以及发展变化的趋势。

综合归纳法是指由个别到一般，由具体事实到抽象概括的推理方法。在统计分析研究中，我们经常将观察到的各种特征归纳得出关于总体的某种信息。例如，平均数概括地反映了总体某一数量标志的一般水平，它与总体中各个个体的标志值不同，但它又是从总体中各个个体的标志值中归纳出来的。通过综合归纳法，可以计算出现象在具体时间、地点条件下的总量规模、集中趋势、离散程度以及分布特征等，并可以进一步从动态上研究现象的发展规律和变化趋势。

3. 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。统计在研究现象的总体数量关系时，需要了解的总体对象的范围往往是很大的，由于经费、时间和精力等各种原因，以致有时在客观上只能从中观察部分单位或有限单位进行计算和分析，根据局部观察结果来推断总体。例如，要说明一批灯泡的平均使用寿命，只能从该批灯泡中抽取一小部分进行检验，推断这一批灯泡的平均使用寿命，并给出这种推断的置信程度。统计推断是现代统计学的基本方法，在统计研究中得到了极为广泛的应用。

4. 统计模型法

模型是以实体、图形或符号等为手段，对真实系统的结构或运动过程的一种表达方式，它是对所研究的真实系统或过程的一种简化、抽象和类比的表示。模型可以分为两类：一是物理模型，它以真实系统的结构和构造作为模型的组成元素，用缩放了尺寸制作与实物系统相似的模型，模型的变量与真实系统的变量完全一样。二是思考模型，它是在认识实体系统之后，根据一定的逻辑变换规则而建立起来的一种刻画系统结构、特征及运动过程的一种表达方式。

统计模型是一种思考模型。它是根据统计资料，运用统计方法，对研究现象的结构和运动过程的一种表达方式，它既是人们认识事