

统计分析概论

(修订版)

李志伟 吴家棉 施家珍 编著

对外经济贸易出版社

统计分析概论

(第五版)

王德胜、周建群、周永强 主编

清华大学出版社

R

统 计 分 析 概 论

(修订版)

李志伟 吴家楹 施家珍 编著

对外贸易教育出版社

统计分析概论（修订版）

李志伟 吴家楹 施家珍 编著

责任编辑 周敏

对外贸易教育出版社出版

新华书店北京发行所发行

北京密云华都印刷厂印刷

开本 850×1168 1/32 • 印张17.25 • 字数463千字

1989年9月 第2版 • 1989年9月 第1次印刷

印数 1—11,000册 • 定价 6.80元

ISBN 7-81000-153-1/F·039

前 言

党的全面开创社会主义现代化建设新局面的纲领，为我国经济建设的各项工作，包括统计工作。指明了前进的方向、目前各财经大专院校正在为加强统计专业教学而增设或改进统计学课程。广大统计人员和财贸职工也大大提高了学习统计专业知识的积极性。本书就是为适应这种形势的发展而编写的，内容着重在介绍统计描述和统计推断的基本理论和方法。这些理论和方法是以数理统计为基础的，应用范围很广泛，但本书主要是针对财经、商业、外贸专业的读者而选题材和事例。

本书共分十三章。除第一章绪论外，第二章至第五章属于统计描述范围，包括统计资料的分类整理、主要统计量数的计算、指数编制和时间序列分析等方面。这些都是对社会经济数据进行表述性分析的最常用的基本方法。这些方法的原理比较简单，对数学水平要求不高，只要懂得算术四则和开方就能运用。我们认为，无论是一个负责经营管理工作的干部或一个基层单位的统计人员，只要能掌握这样一些描述性的分析方法，并能在实际工作中加以运用，就将能明显地加强所在部门工作的定量综合和趋势分析的能力，有助于经营管理水平的提高。

第六章至第八章是概率论的基本理论介绍。第九章至第十三章阐述概率统计的应用，包括参数估计、假设检验、回归及相关分析、统计决策等基本内容。数理统计是数学的一个分支，有严谨的理论体系。财经院校学生或财贸部门职工所从事的具体专业不同，一般说来数学水平都还不太高。作为一本教材，我们希望它能把统计推断的一些基本概念表述清楚，但又要尽量少用复杂的符号和高深的数学推理。要使读者能在有一定理解的基础上运用书中的分析方法。因此本书力求深入浅出，通俗易懂，多用例题来

说明有关论点和计算方法。读者只要具备高中代数知识和一定的推理能力，就不难领会。为了满足各部门有兴趣读者的需要，对大部分公式作了简单的数学证明。其中有些公式的推导比较复杂，则用注解阐述，或者作为附录，附于章末。对数学掌握较少的读者，这些证明可以略过，不致影响对应用方法的了解。如果通过学习，读者对数理统计的一些最基本概念，如随机变量、概率计算、概率分布、抽样分布、中心极限定理等有一定理性认识，对社会经济现象的观察能从概率统计的角度上有所突破，并能初步掌握其应用方法，为进一步统计理论探索和工作实践打下基础，则编写本书的目的就可以说是基本达到了。

本书按每周授课四时计算，一学期可以授完。教师可以针对学生的数学水平和实际需要适当增补或削减部分内容。学习统计分析的目的全在于应用，必须强调课前预习和课外作业，否则基本概念和演算方法就难以真正理解和掌握。书中除第一章外，其余各章均附有习题，学习时应尽可能地多做。对演算习题卷末都附有答案。

本书初版出书后，由于校对上的疏漏，发现不少差错。现值重印之际，由编者重新审校，作了详细的修订，并增补了第十二章。但由于水平所限，立论和编排上的差错，仍在所难免。我们恳切希望读者多加批评指正。

来函请寄

邮政编码:100013

北京 安外 对外经济贸易大学 信箱1号

统计教研室

编 者

1989年1月

目 录

前言

第一章 绪论	(1)
1.1 统计资料和统计学	(1)
1.2 大量观察法是统计科学研究的基本方法	(3)
1.3 统计描述和统计推断	(4)
1.4 统计方法在科学管理中的作用	(6)
1.5 统计研究方法的程序	(7)
第二章 统计资料的搜集、整理和表述	(9)
2.1 统计调查方案的设计	(9)
2.2 统计资料的审查	(15)
2.3 统计资料的汇总	(16)
2.4 统计数列的类型	(17)
2.5 质别资料的分组	(20)
2.6 统计表的一般要求	(22)
2.7 统计图	(25)
2.8 频数分布表的编制方法	(32)
2.9 频数分布的图示	(38)
习 题	(45)
第三章 统计描述的量数	(49)
3.1 相对数	(49)
3.2 频数分布的描述性量数	(52)
3.3 简单算术平均数	(53)
3.4 加权算术平均数	(57)
3.5 中位数	(58)
3.6 众数	(63)

3.7	几何平均数.....	(67)
3.8	离中量数.....	(68)
3.9	方差和标准差.....	(70)
3.10	标准差的意义和应用.....	(73)
3.11	样本的方差和标准差.....	(75)
3.12	变异系数.....	(76)
3.13	偏斜度.....	(77)
	附录: 关于连加号 Σ 的使用方法.....	(78)
	习 题.....	(81)
第四章	指数	(84)
4.1	什么是指数.....	(84)
4.2	简单指数.....	(87)
4.3	加权综合指数.....	(90)
4.4	加权比率平均指数.....	(96)
4.5	指数的基期转移、连锁和编接.....	(100)
4.6	指数的应用.....	(105)
4.7	指数编制的一些问题.....	(112)
	习 题.....	(116)
第五章	时间序列分析	(121)
5.1	时间序列及其组成因素.....	(121)
5.2	时间序列分析.....	(124)
5.3	线性趋势.....	(127)
5.4	非线性趋势.....	(137)
5.5	季节性变动的测量.....	(145)
5.6	周期性变动的测量.....	(157)
	习 题.....	(162)
第六章	概率	(170)
6.1	随机现象和概率模式.....	(170)
6.2	关于概率的基本概念.....	(175)

6.3	概率的加法规则	(178)
6.4	条件概率和联合概率表	(181)
6.5	概率的乘法规则	(186)
6.6	贝叶斯公式	(195)
6.7	事件的计数方法	(202)
	习 题	(210)
第七章	概率分布	(215)
7.1	随机变量	(215)
7.2	随机变量的期望值	(219)
7.3	随机变量的方差	(222)
7.4	随机变量之和的期望值和方差	(228)
7.5	随机变量的典型概率分布	(231)
7.6	二项分布	(232)
7.7	超几何分布	(245)
7.8	泊松分布	(249)
7.9	正态分布	(256)
7.10	标准正态分布及其应用	(261)
7.11	应用标准正态差推算离散型分布的概率	(264)
	附录: 关于公式7.10的证明	(268)
	习 题	(270)
第八章	抽样和抽样分布	(275)
8.1	有关抽样的一些基本概念	(275)
8.2	基本抽样设计	(279)
8.3	样本统计量和抽样分布	(286)
8.4	样本平均数的抽样分布	(288)
8.5	中心极限定理	(293)
8.6	有限总体校正系数	(299)
8.7	比率的抽样分布	(301)
8.8	总体标准差 σ 不明确时的抽样分析	(304)

习 题	(311)
第九章 参数估计	(317)
9.1 点估计和区间估计	(317)
9.2 点估计的评选标准	(318)
9.3 区间估计的基本原理	(321)
9.4 比率的区间估计 (大样本)	(325)
9.5 两个平均数间差别的区间估计 (大样本)	(327)
9.6 两个比率间差别区间估计 (大样本)	(329)
9.7 小样本的区间估计	(330)
9.8 样本量的确定	(333)
习 题	(338)
第十章 假设检验	(341)
10.1 假设检验的基本概念	(341)
10.2 两种类型的误差	(342)
10.3 双尾检验和单尾检验	(343)
10.4 单一样本平均数的假设检验	(347)
10.5 单一样本比率的假设检验	(350)
10.6 第二型误差 β 的计算方法	(353)
10.7 功效曲线	(355)
10.8 样本量和两类误差的关系	(358)
10.9 平均数间差别的检验 (双尾检验)	(360)
10.10 比率间差别的检验	(362)
10.11 小样本的平均数假设检验	(367)
习 题	(368)
第十一章 简单回归分析和相关分析	(371)
11.1 绪论	(371)
11.2 回归模型的形成	(372)
11.3 简单线性回归方程	(374)
11.4 估计标准误差	(382)

11.5	回归分析在统计推断中的应用.....	(387)
11.6	可决系数的意义和计算方法.....	(393)
11.7	相关和相关分析.....	(396)
11.8	对相关系数的推断.....	(402)
11.9	关于回归和相关分析的几个应注意的问题...	(403)
	附录: 关于 Y_c , Y_1 和 b 的方差公式的推导...	(405)
	习 题.....	(408)
第十二章	多元回归分析和相关分析.....	(415)
12.1	包含两个自变量的线性回归方程.....	(416)
12.2	估计标准误差.....	(420)
12.3	多元回归分析中的统计推断.....	(422)
12.4	多个自变量的回归方程.....	(427)
12.5	多元可决系数和相关系数.....	(429)
12.6	偏可决系数和偏相关.....	(431)
12.7	逐步回归法.....	(434)
	习 题.....	(444)
第十三章	统计决策分析.....	(450)
13.1	绪言.....	(450)
13.2	构成统计决策问题的因素.....	(455)
13.3	先验信息条件下的决策分析.....	(455)
13.4	应用树形图的决策分析.....	(457)
13.5	应用后验信息的决策.....	(461)
13.6	关于试验与否的决策.....	(467)
13.7	应用抽样信息的决策.....	(470)
13.8	抽样信息的估值和样本量的选择.....	(475)
13.9	统计决策和传统统计推断的比较.....	(481)
	习 题.....	(484)
	统计附表.....	(491)
	习题答案.....	(504)

词汇索引.....	(523)
公式摘要.....	(531)

第一章 绪 论

1.1 统计资料和统计学

在我们编制国民经济计划、进行企业管理以及从事科学研究的时候，都离不开利用某些数字资料。例如研究我国对外贸易的情况时，我们需要知道全国进出口贸易总额，主要商品的进出口数量。这些数字就是统计资料。下面是一些具体的例子：

(1) 1981年我国进出口总额735.34人民币亿元，出口总额367.61人民币亿元，进口总额367.73人民币亿元。^①

(2) 1980年我国外贸部门出口厂丝7,731吨，绸缎13,225万米。^②

(3) 1980年上海港口吞吐量达到8,400万吨，比1949年增长42倍。^③

统计资料是通过数字在一定空间、时间条件下，反映客观现象的具体数量特征的，而客观现象的数量特征又分为两个方面。一个方面是总数量的多少，如上面我们所举的一些具体例子。另一方面则是客观事物在数量方面的变异 (variability)。

从科学研究的角度来看，仅仅知道客观事物的总数量是不够的。因为从客观事物的组成来看，在个体与个体之间存在着变异。例如1981年根据全国46个城市的抽样调查资料计算，所调查的8,715户，如按平均每人每月生活费收入分组，在20元以下的有179

① 《中国统计年鉴》1981年，第386页。

② 《中国统计年鉴》1981年，第373页。

③ 对外经济贸易大学国际贸易问题所研究《国际贸易问题》1981年，第3期第35页。

户，而60元以上的有566户。^①这是最高和最低的两个组，其所购置的生活用品显然是不同的。收入高的户一般购置高档消费品多，收入低的户则主要购置生活必需品。而我们商品供应既要满足收入低的职工的需要，也要满足收入水平高的职工的需要，究竟按什么比例向市场投放各种商品，这就是一个比较复杂的问题了。又如我们要研究某一个地区的小麦生产情况，仅仅知道这个地区播种了多少亩小麦，总产量有多少万斤，这是不够的，因为分地块来观察，这一块地和另一块地的土壤不同，水利条件不同，所播种的小麦品种不同，播种期有早有晚，施用的肥料种类和数量以及病虫害防治的情况也不相同。这些因素都会影响到产量。所以这一块地和另一块地产量的变异必然是有一定原因的，通过大量观察可以找到比较普遍的规律，为提高单位面积产量，找出科学措施。因此从研究的需要出发，必须有一套科学方法以分析大量现象的数量变异，从中找出客观规律，这种科学方法就是统计学所提供的方法。所以统计学是对大量统计数字进行分析研究的科学方法。

数学是一切科学研究数量关系的工具。现代统计学亦即数理统计学是应用数学的一个分支。而本书则着重于讨论统计方法在经济领域中的作用。

“统计”一词在我国古代汉语中仅为“统而计之”之意。^②现代汉语中“统计资料”、“统计工作”和“统计学”是三个涵义不同的名词，在不会混淆涵义的情况下，口语常简称为“统计”，但这并不能理解为“统计”一词有三种涵义。英文中 statistics

① 《中国统计年鉴》1981年，第428页。

② 《清文献通考》“凡天下之赋，亦曰徭里银。顺治十八年统计，直省徭里银三百万八千九百五两有奇，米二万一千五百七十石有奇”。按顺治十八年为1662年，清代直省即直隶省，今河北省。这是我国古汉语中统计二字连用之一例。

一词，作为复数词应用时，指的是统计资料，作为单数词应用时，则指统计学。统计资料是反映客观事物数量特征的数字资料，并不是统计学的研究对象，而统计学则是一门属于方法论的科学。

1.2 大量观察法是统计科学研究的基本方法

在上节中我们已经举例说明了客观事物中个体与个体之间存在着变异，这种变异是普遍存在的。例如在同一民族、同一性别、同一年龄的人群中，其身高、体重是不相同的；同一个学校的同年级的学生某一门学科的学习成绩也不会完全一样。产生以上种种变异的原因是多种多样的。以人的身高为例，有先天的遗传的因素，有后天的营养因素，体育锻炼因素和卫生保健因素等等。但是如果我们观察的数量相当大，比如测量100名北京市18岁男子的身高，就可以得到一个可以说明这个年龄男性的平均高度，其中由于某些造成特殊高的或特别低的因素的影响一般会在大量观察中被冲消，因而显示出一般的规律性来。按同样的道理，我们研究某一个地区的小麦生产水平，如果不是仅仅观察个别地块的产量，而是观察大量地块的产量，则某些个别地块由于特殊原因而引起的特殊高产或特殊低产，一般可以相互抵消，而显示出代表该地区一般生产力的平均水平。这种从大量观察以求得表现一般规律的科学方法是统计学的基本方法。

客观集体大量现象的数量差异乃是必然性与偶然性，或共性与个性的对立统一规律的具体表现。马克思主义哲学的辩证唯物主义，对于应用统计学（也即数理统计方法的应用）在理论上是具有指导意义的。如上面所举的例子，个别人身材特高或特低，或个别地块的产量特别高或特别低都是偶然性的表现。正如革命导师恩格斯所说：“在表面上是偶然性在起作用的地方，这种偶然性始终是受内部的隐蔽着的规律支配的，而问题只是在于发现

这些规律。”^①

虽然客观现象是非常复杂的，但是经过分析可以把复杂的现象加以简化，找出明确的规律来。例如生丝是我国传统的出口商品，我们要保持我国生丝在国际市场上的竞争能力，必须在保证质量优异的前提下，不断降低生产成本，以及各项包装储运成本等等。这些成本当然不是调查一两个生产单位所能决定的，必须调查相当大量的生产单位，以求得平均成本。尽管实际工作非常复杂，但从方法上来看，仍是通过大量观察以取得平均数的问题。当然问题的研究不能到此为止，还要通过分析比较，找出降低成本的途径来。

大量观察法或平均法的应用，不仅限于取得一般意义的平均数，在分析大量观察的数量变异上还有许多用途。例如在分析数量变异方面常用的标准差（在本书第三章论述），分析两个或两个以上变量的相关关系等（在本书第十一、十二章论述），都是大量观察法的推广应用。所以大量观察法是统计科学研究的基本方法。

1.3 统计描述和统计推断

统计学的内容主要包括两个方面，即描述统计学和归纳统计学。

描述统计学（descriptive statistics）主要包括对已收集到的统计资料如何进行整理（如按地区分组、按时间序列整理、按质别分类、按数量分组等）、编制统计表、绘制统计图、计算必要的相对数、平均数、标准差、编制指数、描述时间数列资料的长期趋势与季节变动，测定两个或两个以上的现象的相关关系等等。这些工作总的目的，就是使反映客观现象的统计数字可以一目了然，条理清晰，便于使用。本书的前五章，以及第十一、

^① （马克思恩格斯选集）中译本第4卷，第234页。

十二章（回归分析和相关分析）的前半部分主要讨论这个问题。

归纳统计学（inductive statistics），也称为统计推断（statistical inference），是统计学的一个重要组成部分，其主要内容是研究如何应用概率理论，从样本以推论总体。

所谓总体（population）是指特定的统计研究对象的全部个体的总和。例如，我们研究北京市职工家庭收入，则北京市全部职工家庭的家庭收入就是总体。从总体中抽出来的若干户作为调查对象，这若干户的家庭收入就是样本（sample）。如果要对北京市全部职工家庭逐户调查其收入情况，在实际工作中是行不通的，而且也没有这种必要。因此只能抽取若干户职工家庭（例如抽取2,000户）进行调查，从样本调查的结果以推测总体的某种特征。

从样本调查结果以推论总体，不仅限于推论总体的平均数或比率（例如在一大批产品中合格产品所占的百分比），也包括假设检验，这在科学研究中是一个非常重要的方法。例如在农业生产中可以检验新品种和对照品种是否存在着显著性差异，以判断新品种是否有明显的增产效果，有无推广的价值。在工业生产上可以用于检验两种不同的工艺方法所生产的产品在质量上有无显著的差别，从而判断一种新的工艺方法是否优于原有的工艺方法。在进出口商品检验中，也常用到这种方法。例如我们进口（或出口）一批工具钢，按照合同规定这种钢材必须耐高温 600°C 。而抽取100个样品检验的结果，样品耐高温 590°C ，在这种情况下买卖双方可能发生争议，一方认为这种钢材不合格，而另一方则认为全部钢材总的来说是合格的，样品检验结果虽未达到合同所规定的标准，但可能是由于抽样的偶然性所引起。究竟双方的意见以何者为合理，可以通过统计学的假设检验加以确定。

统计推断是以概率论为基础的，为了便于读者学习，我们在第六章和第七章论述了有关概率论的基础知识，第八章讨论抽样和抽样分布，这都是为第九章参数估计，第十章假设检验，第十一、