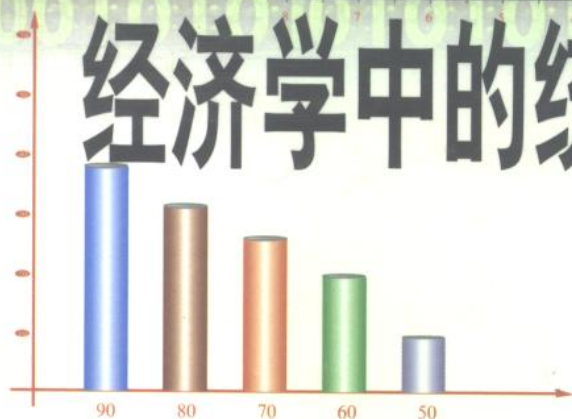


高等院校选用教材系列

经济学中的统计方法



王学仁 编著



科学出版社

Science Press

高等院校选用教材系列

经济学中的统计方法

王学仁 编著

科学出版社

2000

内 容 简 介

本书对经济研究必需的统计概念和方法作了一系列的介绍,从实用的角度出发,尽量做到经济学和数理统计学的结合,突出经济应用的背景,显示统计计算在财经方面的重要性,并介绍了一些对经济研究十分有用的新的统计方法。

本书可作为统计学专业及经济、管理类专业的本科生和研究生的教材;对于经济学家、企业管理专家及财经工作者也有阅读和参考价值。

图书在版编目(CIP)数据

经济学中的统计方法/王学仁编著. -北京:科学出版社, 2000.5

ISBN 7-03-007771-7

I. 经… II. 王… III. 经济统计-方法 IV. F222.1

中国版本图书馆 CIP 数据核字 (1999) 第 31706 号

2/38/12

科学出版社 出版

北京东黄城根北街16号
邮政编码:100717

新蕾印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

2000年5月第 一 版 开本: 850×1168 1/32

2000年5月第一次印刷 印张: 13 1/4

印数: 1—3 000 字数: 347 000

定价: 21.00 元

(如有印装质量问题,我社负责调换〈北燕〉)

前 言

由于电子计算机的广泛应用,促进了工商管理和一切财经活动数量化的要求。如何搜集、整理和分析经济的资料和数据,以揭示其内在的规律并预测其进一步发展的趋势,已经提到广大经济工作者和企业家的面前。这种从定性到定量的变化,要求经济学家、企业管理专家不仅要有精深的经济管理知识,还要能运用计算机和掌握数理统计学的概念和方法,并能把经济、统计、计算机三者有机地结合,以经济问题为对象,用数理统计的方法去研究、处理和分析经济数据,通过计算机计算,提出经济决策的数量依据。三者的结合对于培养青年经济工作者尤为重要。

本书的目的:对经济研究必需的统计概念和方法作一系列的介绍。从实用的角度出发,尽量做到经济学和数理统计学的结合。在取材上,突出经济应用的背景,显示统计计算在财经方面的重要性,并包括一些对经济研究十分有用的新的统计方法。在书中,对数学上的推导和计算过程,尽量采用简明易懂的方法处理。读者只要具有最基本的数理统计知识即可读懂。本书编写过程中,参阅了 [3]、[4] 和 [5] 中的有关内容和实例,特此表示感谢。

现在,已经有“生物统计”、“医学统计”、“地质统计”、“工业统计”、“农业统计”、“社会统计”,以及“经济统计”等许多统计书藉,说明对数据的研究和定量分析已经引起各方面的重视。希望本书的出版能激励广大经济工作者,注意运用数学的重要性,更望能抛砖引玉,使将来有更多这方面的好书出现。限于学识和经济实践的经验,本书仍有许多不足和值得商榷的地方,恳请广大读者批评指正。

本书是在美国欧伯林大学编写完成的。欧伯林大学和欧伯林山西协会为本书的编写提供了许多条件。特别地,欧伯林山西协会主任卡尔·贾克忱博士为本书的编写提出了许多宝贵的意见,

帮助查找资料，在各方面给予很多帮助，特此表示感谢。为增强云南大学和欧柏林大学多年的合作和友谊，仅以此书献给欧柏林大学和欧柏林山西协会。

编 者

目 录

第一章 统计的基本概念和经济数据的整理	1
§ 1.1 引言	1
§ 1.2 统计的基本概念	2
§ 1.3 经济资料的整理——样本的图形表示法	4
§ 1.4 经济资料的频数分布和直方图	9
§ 1.5 经济资料的分析——样本的表征数	13
1.5.1 集中趋势的计量指标	13
1.5.2 平均数、中位数和众数的关系	18
1.5.3 离散趋势的计量指标	18
1.5.4 偏态的计量	23
§ 1.6 矩阵及其应用	24
1.6.1 向量	25
1.6.2 矩阵	27
1.6.3 行列式	30
1.6.4 逆矩阵及其运算	31
1.6.5 二次型和正定矩阵	34
1.6.6 剖分向量和矩阵	35
1.6.7 有关矩阵求导数的定义和几个定理	39
1.6.8 矩阵应用的例子	40
§ 1.7 数据处理	45
1.7.1 数据的标准化处理	45
1.7.2 数据的转换	48
1.7.3 正态性检验	50
1.7.4 偏度检验 ($8 \leq N \leq 5\,000$)	55
1.7.5 峰度检验 ($7 \leq N \leq 5\,000$)	56
1.7.6 偏度和峰度联合检验(多方向检验 $20 \leq N \leq 1\,000$)	58
第二章 统计推断	70
§ 2.1 假设检验	70

2.1.1	假设检验的方法	70
2.1.2	备择假设	72
2.1.3	两类错误	73
2.1.4	对正态总体平均值的检验	79
2.1.5	对正态总体方差的检验	81
2.1.6	总体分布的鉴定——拟合优度检验	83
§ 2.2	参数估计	86
2.2.1	估计量的性质	86
2.2.2	构造估计量的方法	99
2.2.3	区间估计	111
第三章	一元回归分析	116
§ 3.1	一元回归分析的模型	116
§ 3.2	回归系数的最小二乘估计	120
§ 3.3	回归估计的统计推断	123
3.3.1	$\hat{\alpha}$ 和 $\hat{\beta}$ 的分布	123
3.3.2	$\hat{\alpha}$ 和 $\hat{\beta}$ 的协方差	124
3.3.3	σ^2 的估计式	126
3.3.4	半平均数法拟合回归直线	129
3.3.5	α 和 β 的区间估计	131
3.3.6	$E(y_i)$ 的区间估计	133
3.3.7	y 的样本变差的分解	136
3.3.8	回归方程的假设检验	140
3.3.9	回归分析的表述	142
§ 3.4	预测和控制	142
3.4.1	预测	142
3.4.2	控制	145
3.4.3	实验的策略	147
§ 3.5	化曲线为直线的回归分析	150
第四章	基本假设被违反的回归估计	154
§ 4.1	异方差性	155
4.1.1	异方差情况下最小二乘估计的性质	155
4.1.2	最小二乘估计式估计方差的性质	156
4.1.3	关于 σ_i^2 的设定	158

4.1.4 关于异方差性的检验	167
§ 4.2 自回归扰动	169
4.2.1 自回归扰动的生成模式	170
4.2.2 最小二乘估计量的性质	172
4.2.3 一阶自回归模型的极大似然估计	177
4.2.4 二段迭代估计	181
4.2.5 德宾(Durbin)法	182
4.2.6 一阶差分的应用	183
4.2.7 非自回归检验	186
§ 4.3 随机的解释变量	188
第五章 不完备数据的回归估计	197
§ 5.1 存在测量误差的回归估计	197
5.1.1 工具变量估计	198
5.1.2 方程误差模型	203
5.1.3 变量误差模型	204
5.1.4 组平均法	204
5.1.5 加权回归	206
§ 5.2 分组数据的回归估计	210
5.2.1 单向分组数据的回归估计	210
5.2.2 双向分组数据的回归估计	217
§ 5.3 缺失数据的回归估计	224
5.3.1 非随机解释变量的缺失	224
5.3.2 零回归估计	228
第六章 多元回归模型	231
§ 6.1 回归参数的最小二乘估计	231
§ 6.2 多元回归方程的估计和检验	235
6.2.1 最小二乘估计量的方差和协方差	235
6.2.2 σ^2 的估计	239
6.2.3 $E(y_i)$ 的置信区间	242
6.2.4 y 的样本变差的分解	243
6.2.5 多元回归方程的假设检验	245
6.2.6 预测	253
6.2.7 计量单位的变换	255

6.2.8 关于多元回归模型基本假设和数据不完备的注释	258
§ 6.3 多元共线性	259
6.3.1 多元共线性	259
6.3.2 完全多元共线性	260
6.3.3 高度多元共线性	265
6.3.4 多元共线性的测量	267
§ 6.4 设定误差	268
6.4.1 忽略一个有关解释变量的误差	268
6.4.2 解释变量性质上的变化	272
6.4.3 包含一个无关解释变量的误差	272
6.4.4 非线性性	275
6.4.5 扰动项的不正确设定	276
6.4.6 设定误差检验	278
第七章 模型及其估计	281
§ 7.1 二态变量模型	281
7.1.1 单一二态解释变量	281
7.1.2 几个定性解释变量	284
7.1.3 交互作用	287
7.1.4 定性和定量的解释变量	288
7.1.5 二态因变量	294
7.1.6 定性因变量和定性自变量	296
§ 7.2 约束系数模型	298
7.2.1 定期约束	298
7.2.2 不等式约束	300
7.2.3 线性约束	305
7.2.4 非线性约束	308
7.2.5 约束对 R^2 和预测误差方差的影响	313
7.2.6 约束在估计中的作用	314
§ 7.3 非线性模型	315
7.3.1 内蕴线性模型	315
7.3.2 内蕴非线性模型	323
7.3.3 线性性检验	328
§ 7.4 分布滞后模型	333

7.4.1 几何滞后	334
7.4.2 帕斯卡(Pascal)滞后	346
7.4.3 多项式滞后	350
第八章 联立方程组	353
§ 8.1 联立方程组的描述	354
8.1.1 联立模型的结构形式	354
8.1.2 联立模型类型	360
§ 8.2 识别问题	362
8.2.1 确定方程的识别	362
8.2.2 约束干扰的方差-协方差矩阵的识别	368
8.2.3 不足识别	370
§ 8.3 联立方程中单一方程的估计	371
8.3.1 精确识别方程的估计	372
8.3.2 二段最小平方估计	380
8.3.3 K 阶估计	387
8.3.4 有限信息极大似然估计	389
§ 8.4 方程组的系统估计法	395
8.4.1 三步最小平方估计	396
8.4.2 完全信息极大似然估计	401
§ 8.5 动态经济模型分析	404
8.5.1 一个简单的经济模型	404
8.5.2 稳定性条件和动态分析	407
参考文献	411

第一章 统计的基本概念和经济数据的整理

§ 1.1 引言

近代的经济活动中,我们经常会接触到许多数据资料,例如一个国家(或地区)在一定时期内全体居民的收入,几个时期内一个国家(或地区)的国民收入,工厂产品的产量与市场销售,家庭的收入与支出,商品的成本、售价和盈利,以及市场需求弹性预测,市场货币需求量预测,转口贸易预测,股票市场预测等等.在这些数据资料中往往包含着许多重要而有用的信息.一般来说,从各方面所汇集的资料与情报,其中的规律性并非一目了然,而是蕴藏在大量数据之中,因此必须应用数理统计的方法把数据加以重新整理,进行系统的分析和研究,揭示数据的统计规律性,正确地提取有用的情报信息,以便为制订经济策略,提供可靠的数量依据.

统计学的对象是数据.它是研究如何搜集、整理和分析数据的科学,随着经济的繁荣,在经济活动中要求数量化方法的不断增加,统计学应用的范围也就越来越广泛.概率论的发展又带来了统计学理论的巨大进步,从而可以对不完整的信息进行科学的概括,即统计推断.结果导致“统计”学科由“描述统计学”与“推断统计学”两部分组成.描述统计学阐述资料的搜集,数据的整理和分析.统计推断学则讨论怎样由部分概括去推断整体.

鉴于在商业和经济活动中,最基本的统计概念、知识和方法、统计图表的应用具有广泛而重要的作用;而统计学的理论几乎是计量经济学的基础.因此,在本章中我们要介绍几个基本的经济概念、统计图表的绘制与应用、频数分布、特征量以及一些有用的数据处理方法.同时还要介绍作为计量经济学理论基础的几个最重要的统计推断方法.

§ 1.2 统计的基本概念

样本和总体 我们经常用到的经济观测数据称为样本或称为子样,是相对一定的总体而言的.

总体是某次统计分析工作中所欲研究对象的全体,总体又称母体.对于总体中的每一个元素称为个体.因之,总体是被观察变量所有可能观察结果的总和(研究对象的某种数量性质的一个数值集合,或者说,总体是某个随机变量 X 取值的全体),例如一个国家一定时期内全体居民的收入;几个时期内一个国家的国民收入;以及重复地观察某医院出生婴儿性别的全部可能结果等等.总体中所包括个体的数目叫做总体的容量.当一个总体是由有限个个体(或数值)组成时,称为有限总体,而当一个总体由无限个个体组成时,就称为无限总体.

与总体相关的概念是样本.样本是用以推断总体的性质而从中抽取出来的一部分个体的集合.样本中的每个个体称为样品.显然,总体是样本的全体,它们的关系是部分与全体的关系,往往总体是未知的,样本是已知的.但是,要注意样本是随机抽取而得到的(总体中的每一个体被抽取的机会是相等的),在抽取之前,它取什么值是随机的,即样本实现哪一组值事前是无法知道的.例如,研究某地区在一定时期内家庭的人均收入,我们随意选取本地区 50 个家庭进行调查,那么,在抽察之前它们都可以从整个地区在一定时期内全部家庭收入中取到任何一组值(逐次抽取),所以,每次抽样值可以看成是一个独立的随机变量,50 次抽样组成一组相互独立的随机样本(即从总体中可重复的 n 次随机抽样): $x_1, x_2, \dots, x_{50}$. 但一经观测以后,其实现值就组成一组具体的样本值了.所以,样本具有两重性,一方面它们在观察以前到底取总体中哪一组值是随机的,它们是一组相互独立的随机变量($x_1, x_2, \dots, x_n$);另一方面,在一次具体试验中,经取样观察以后,所获得的是一组确定的观察值,称为样本值: $x_1, x_2, \dots, x_n$. 以后,我们统计分析方

法就是从样本值出发推算出有关结果来。

在本节我们要介绍的最后一个概念是分布. 对样本来说有频数分布, 而对总体来说则有概率分布, 诸如总体是服从正态分布、 t -分布、 F -分布或 χ^2 -分布等等. 这些常用的分布在一般数理统计的教科书中都有详细的介绍, 我们这里就不重复赘述了. 关于样本的频数分布, 一种常用的方法是对抽取数据加以整理, 作直方图, 画出频率分布曲线, 这就可以直观地看到数据分布的情况, 在什么范围, 较大较小的各有多少, 在哪些地方分布得比较集中, 以及分布图形是否对称等等. 所以样本的频数分布是总体概率分布的近似. 有时, 还可以通过样本作出样本的分布函数 $F^*(x)$.

设 $(x_1, x_2, \dots, x_n)$ 为从总体中随机抽取的一个容量为 n 样本, 令

$$F^*(x) = \frac{k}{n},$$

其中 k 是 $(x_1, x_2, \dots, x_n)$ 中小于 x 的个数, 显然, 对于固定的 x , $F^*(x)$ 是 $(x_1, x_2, \dots, x_n)$ 的函数, 又称它为经验分布函数. 特别当 n 很大时 $F^*(x)$ 与总体的分布函数 $F(x)$ 是十分近似的, 所以常常可以从样本推测出总体的分布来.

在统计分析中, 对总体提出的问题是多种多样的, 有时要求了解总体的全貌, 有时要求估计总体的某些特征, 这就要针对不同的要求, 构造出样本 $(x_1, x_2, \dots, x_n)$ 的一些数字特征, 以便推断总体的表征值.

对于样本 $(x_1, x_2, \dots, x_n)$ 其数字表征很多, 诸如通常所熟知的平均值、中位数、众数及方差、极差、变异系数等等, 但其中最常用的是平均值

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

以及样本方差

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

在实际工作中分别用它们来作为总体期望值(数学期望)和方差的

估计量,至于如何进行将在下面参数估计时详细讨论.

值得注意的是,由于样本有两重性,所以样本表征数也有两重性.一方面作为样本 $(x_1, x_2, \dots, x_n)$ 的函数它也是一个随机变量,对不同的试验其取值是不同的.另一方面,在试验之后(对不确定的样本值)表征数就是完全确定的数值.在试验前,前者是主要的,这时样本和样本表征数都是随机的;在试验后,后者是主要的,此时样本值和表征数就是完全确定的了.在分析评价一个统计方法时,因为要考虑到运用这一方法可能产生的各种后果,更多地把样本和表征数看成随机变量.而在具体运用一个统计方法时,就是从样本值出发来计算表征数的.所以掌握好统计中这一矛盾对于深入领会统计学的基本概念是很重要的.以后,我们对于样本 $(x_1, x_2, \dots, x_n)$ 的研究就是从这两重性出发的.

下面,我们将进一步介绍统计资料的整理分析,以及统计数据处理的常用方法^[4].

§ 1.3 经济资料的整理——样本的图形表示法

图形可以给人以直观明了的形象.我们所搜集到的经济和工商业的大量资料大都可以用图形表示的方法加以展示,显现其变化的规律性,从而使人们对这批资料获得全面的了解和正确的解释.

下面给出几种常用的数据表示图.

1 折线图 如果我们研究的对象包括两个变量,这时,可将两变量相对应的各对数值在坐标平面上描点,然后用直线把相继两点联接起来,便得折线图.这种图形可用于显示出两个不同变量间的关系,特别表现一个变量在一段时期内所发生的变动情况.例如,根据统计资料的记载,某个商业银行的存款从1965~1990年(每隔5年)的总存款(亿元)列于表1.1中.

为了研究在1965~1990年时期内商业贸易发展的情况,可将上述资料画成折线图如图1.1所示.

表 1.1

年 份	1965	1970	1975	1980	1985	1990
存款总额(亿)	149.9	156.1	193.2	230.5	333.8	485.5

2 曲线图 在折线图中,如果联结各点的不是一串直线,而是一条光滑曲线,即得曲线图,随着描点的增多,曲线图和折线图几乎完全趋于一致,曲线图适宜于表示连续变动的变量,如气温、雨量、谷物收获量、银行储蓄额、物价等。这两种图都可在各个值所描绘的区间内,从一个变量的已给数值估计另一个变量数值,而在描

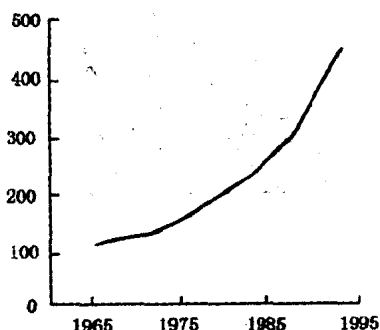


图 1.1 折线图

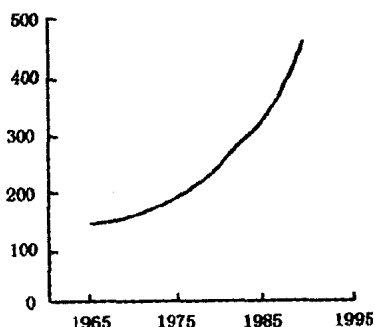


图 1.2 曲线图

绘区间外的估计值,往往是不可靠的。

折线图比较简单,容易绘制,但在某些问题中,曲线图表示描点间的变动要真实些,而这两种图形在图示两变量间关系时,是同样有效的(图 1.2 表示表 1.1 数据的曲线图,从图上估计 1983 年商业银行存款总额约计 300 亿元。

3 条形图 条形图是用平行长条作比较的,画成水平式或垂直式两种条形图。条形图适用于展示一段时期内单一的统计资料数列。在条形图中,各长条的宽度与各长条间的距离彼此均等,条的长度与所代表的变值成比例。

图 1.3 描绘机械制造业工人每小时创造的总值的垂直条形图;图 1.4 表明各部门工人每小时生产总值的比较。

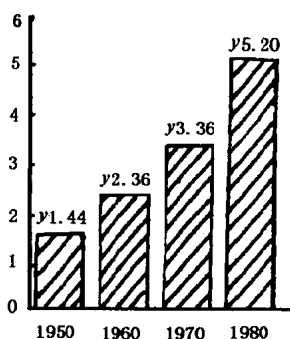


图 1.3

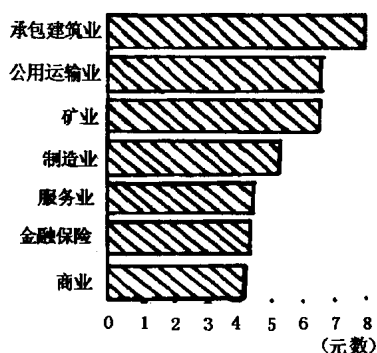


图 1.4

4 合成条形图 这是将总量的变动与其组成部分变动对比的图示法。如果要对比各组成部分,而不把它们同总量联系起来,则合成条形图是一种有效的方法。

例 1.1 某百货商店设有 3 个售货部:领带部、鞋袜部和服装部。经营管理需要用图表示 1976、1977、1978 这 3 年中总销售额的变化和各售货部销售额占总额比例的变化。兹将 3 个部的销售额分别列于表 1.2。

表 1.2

部 门	销售额(百万元)		
	1976	1977	1978
领 带	2	3	2
鞋 袜	10	8	3
服 装	4	8	18
总 额	16	19	23

为了绘制一个合成条形图,分别画出 1976 年、1977 年、1978 年各部门销售额叠加图,然后合成条形图(见图 1.5)。由合成条形图可以看出:

(1) 3 年总的销售额是增长的;

- (2) 3年内领带的销售额保持相对稳定;
- (3) 作为总体的组成部分,鞋袜的销售额减少了;
- (4) 作为总体的组成部分,服装的销售额增长迅速.

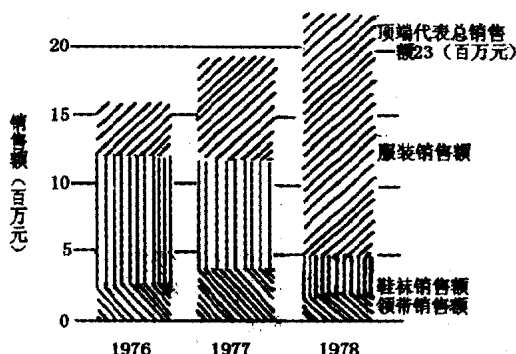


图 1.5 合成条形图

5 双向条形图 这种图可用来表示收益和亏损,超过正常或低于正常的活动以及从一个时期到另一个时期百分比的变化.

例 1.2 把某商店 1978 年头 6 个月的销售额与 1977 年头 6 个月的销售额进行比较,目的在于说明销售额百分比的变化,而不是金额数量的变化.其销售额资料列表 1.3 中.

表 1.3

部 门	头 6 个月的销售额(元)		1977 年到 1978 年变动的百分比
	1977	1978	
腰 带	4 000	3 000	-25
金属器具	1 000	1 500	50
收音机	10 000	5 000	-50
电视机	100 000	110 000	10
小器械	25 000	50 000	100

$$1977 \text{ 年到 } 1978 \text{ 年变动百分比} = \frac{1978 \text{ 年销售额} - 1977 \text{ 年销售额}}{1977 \text{ 年销售额}} \times 100\%.$$