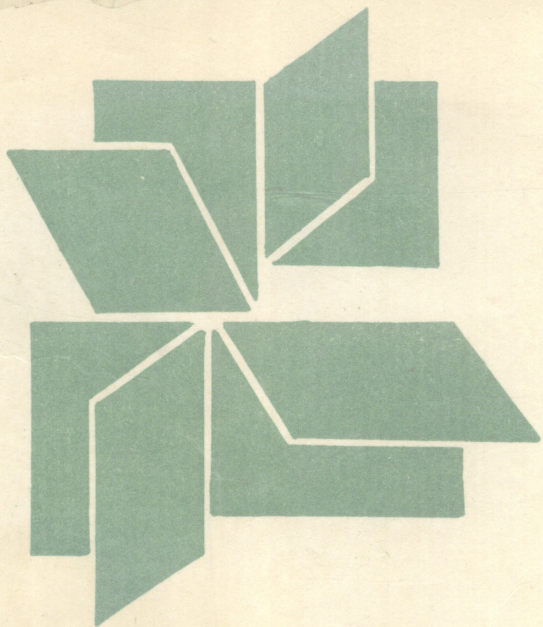


全国高等教育自学考试教材

JINGJIGUANLILEIZHUYANYEJIAOCAI

经济管理类专业



主编 唐国兴

高等数学(二)  
第二分册 概率统计

武汉大学出版社

全国高等教育自学考试教材

# 高等数学

(二)

第二分册 概率统计

(经济管理类专业)

唐国兴 主编

武汉大学出版社

1991·武昌

## 图书在版编目(CIP)数据

高等数学 第二分册:概率统计/唐国兴著

——武汉:武汉大学出版社,1991.1

全国高等教育自学考试教材

经济管理类专业用

ISBN 7-307-00914-5

I. 高…

Ⅱ. 唐…

Ⅲ. ①概率统计—教材 ②高等数学—多卷书—教材

Ⅳ. O13

武汉大学出版社出版发行

(430072 武昌 珞珈山)

湖北崇阳县印刷厂印刷

1991年1月第1版 1995年11月第13次印刷

开本:850×1168 1/32 印张:13.5

字数:343千字 印数:205501—235500

ISBN 7-307-00914-5/O·83 定价:14.30元

## 出版前言

高等教育自学考试教材建设是高等教育自学考试工作的一项基本建设。经国家教育委员会同意,我们拟有计划、有步骤地组织编写一些高等教育自学考试教材,以满足社会自学和适应考试的需要。《高等数学(二)概率统计分册》是为高等教育自学考试经济管理类专业组编的一套教材中的一种。这本教材根据专业考试计划,从造就和选拔人才的需要出发,按照全国颁布的《高等数学(二)自学考试大纲》的要求,结合自学考试的特点,组织高等院校一些专家学者集体编写而成的。

经济管理类专业《高等数学(二)概率统计分册》自学考试教材,是供个人自学、社会助学和国家考试使用的。无疑也适用于其他相同专业方面的学习需要。现经审定同意予以出版发行。我们相信,随着高教自学考试教材的陆续出版,必将对我国高等教育事业的发展,保证自学考试的质量起到积极的促进作用。

编写高等教育自学考试教材是一种新的尝试,希望得到社会各方面的关怀和支持,使它在使用中不断提高和日臻完善。

全国高等教育自学考试指导委员会

一九九〇年七月

## 序 言

本书是按照经济管理类大学本科高等数学(二)自学考试大纲编写的。

在经济科学和管理科学中,我们经常碰到诸如抽样调查、预测、决策一类的问题,在这些问题中,我们所研究的对象常常具有随机性。概率统计是对随机现象的统计规律进行演绎和归纳的科学,因此它在经济和管理中有着广泛的应用。学好本课程将有助于读者对现代经济管理科学中许多理论和方法的理解与掌握。

本书力求以通俗易懂的语言向读者介绍概率统计的基本概念、基本理论和基本方法。全书共分九章。第一章介绍描述统计方法,目的是使读者了解对统计数据的简单归纳方法。第二章介绍概率论的最基本概念,即随机现象及其描述的方法。第三章引进了随机变量的概念,并介绍了用分布函数描述随机变量统计特性的方法。第四章可以认为是概率论和数理统计的连接界面,介绍了抽样和抽样分布以及大数法则和中心极限定理。第五章讨论了如何利用随机子样来估计总体参数的方法,并提出了评价估计量优良性的若干标准。第六章介绍如何利用子样来对总体的特征进行检验的方法。在第七章中,我们用较多的篇幅介绍了概率统计在产品质量管理中的应用。第八章讨论了对反映随机变量之间因果关系的回归模型进行估计和检验的方法。在第九章中,我们进一步介绍了概率统计在经济和管理中的一些应用、包括经济预测和风险型决策等。考虑到读者自学的特点我们在文中适当的地方插入了一些思考题,以帮助读者进行自我测定。对于打\*号的章节,初学时可以跳过,也不作考试要求。本书各章后附有习题,对于较难的习题,





# 目 录

第一章 描述统计 .....	(1)
§ 1.1 统计资料的来源和整理 .....	(2)
§ 1.2 图形描述 .....	(4)
§ 1.3 位置特征 .....	(19)
§ 1.4 变异特征 .....	(28)
本章要点 .....	(32)
习题 .....	(33)
第二章 概率的基本概念 .....	(35)
§ 2.1 事件及其概率 .....	(35)
§ 2.2 古典概型 .....	(46)
§ 2.3 概率的基本性质 .....	(51)
§ 2.4 条件概率 .....	(54)
§ 2.5 独立重复试验 .....	(66)
本章要点 .....	(69)
习题 .....	(70)
第三章 随机变量与概率分布 .....	(74)
§ 3.1 随机变量 .....	(74)
§ 3.2 离散型随机变量 .....	(78)
§ 3.3 连续型随机变量 .....	(86)
§ 3.4 随机变量的数字特征 .....	(101)
§ 3.5 二维随机向量 .....	(115)
本章要点 .....	(137)
习题 .....	(138)

第四章 抽样和抽样分布 .....	(142)
§ 4.1 随机抽样 .....	(142)
§ 4.2 大数定律和中心极限定理 .....	(150)
§ 4.3 抽样分布 .....	(154)
本章要点 .....	(160)
第五章 参数估计 .....	(161)
§ 5.1 参数的点估计 .....	(162)
§ 5.2 估计量优良性的标准 .....	(173)
§ 5.3 参数的区间估计 .....	(179)
本章要点 .....	(192)
习题 .....	(192)
第六章 假设检验 .....	(196)
§ 6.1 假设检验问题 .....	(196)
§ 6.2 假设检验的程序 .....	(201)
§ 6.3 关于正态总体的假设检验 .....	(204)
§ 6.4 概率的假设检验 .....	(211)
§ 6.5 两个正态总体的比较 .....	(213)
§ 6.6 假设检验的两类错误 .....	(220)
§ 6.7 分布函数的拟合优度检验 .....	(228)
本章要点 .....	(238)
习题 .....	(239)
第七章 工序质量控制和抽样检验 .....	(244)
§ 7.1 工序质量控制 .....	(245)
§ 7.2 计数抽样检验 .....	(261)
本章要点 .....	(274)
习题 .....	(275)



第八章 回归分析与相关分析 .....	(277)
§ 8.1 一元线性回归 .....	(278)
§ 8.2 相关分析 .....	(308)
§ 8.3 一元非线性回归 .....	(316)
§ 8.4 多元线性回归 .....	(321)
本章要点 .....	(327)
习题 .....	(328)
第九章 经济预测与决策 .....	(331)
§ 9.1 因果关系预测 .....	(333)
§ 9.2 简单的时间序列预测 .....	(340)
§ 9.3 风险型决策 .....	(365)
本章要点 .....	(375)
习题 .....	(376)
附表 .....	(379)
参考书目 .....	(411)
习题答案 .....	(412)
后记 .....	(419)

## 第一章 描述统计

翻开“中国统计年鉴”，你可能首先会被里面的许多统计图所吸引，这些图系统地反映了1949年以来我国社会经济各方面的发展情况。例如：在1984年版的年鉴中，在国民收入发展速度的统计图上，可以看出，从1952年起的31年中，我国的国民收入大约增加了5.4倍，但是在1962年和1968年，增长曲线上出现了两个巨大的低谷。在人口年龄构成图所描述的一个金字塔，表明我国的人口结构是属于“年轻型”的或“发展型”的，它同一些西方国家，如日本、瑞典、美国的人口年龄结构大不相同。

类似上述的统计图，我们会经常碰到。工厂的车间里，挂着各班组的出勤、安全生产、产品质量情况的统计图，它们是企业管理工作的一部分；在公安局交通队的办公室里，挂着本地区各月份交通事故数目和事故种类的统计图，它提醒人们必须采取必要的预防措施。这些用图形描述统计数字的方法是传递信息、指导决策的一种重要手段，是统计学的重要组成部分。

一般说来，描述统计包括图形描述和数字特征描述两大类。前者是用各种形象直观的图形来反映统计数据的特征。用什么样的图形以及怎样画统计图，这本身也是一门学问，称为统计制图学。数字特征描述是对一组统计数据作适当的计算处理后，求出反映这组数据特征的一些“量”的方法。例如：要比较北京和上海的职工收入情况，若对具体的个人进行比较，则无法得出有意义的结论，而若采用“平均工资”这一“特征”，1984年北京全民所有制单位职工平均工资为1127元，而上海的为1160元，从而我们比较能够接受“上海职工的收入比北京的高”这一结论。

## § 1.1 统计资料的来源和整理

统计资料来源于对事物的大量观测. 1985 年我国社会总产值为 16 309 亿元, 这一数据只是 1986 年版“中国统计年鉴”所刊载的庞大数据量中的一个, 但它是大量统计工作的产物, 我们必须从每户农民所收获的粮食、棉花、蔬菜以及饲养的家禽家畜、种植的树林等等的价值算起, 从每个工业企业的各种产品的产值算起, 从每天的铁路、公路、水运的报表开始. 我国的农户有 1 亿 8 千多万, 工业企业有 50 万个, 由此我们可以体会到在 16 309 亿元这个数字后面所进行的“大量观测”的含义了.

当然, 对大多数研究者来说, 有各种现成的统计资料可被利用. 目前我国已有多种统计书刊, 如“中国统计年鉴”、“海关统计”(季刊)、“中国统计月报”等等, 从联合国(U. N.)和国际货币基金组织(IMF)等出版的统计杂志也能查阅到有关的数据. 但是, 我们在研究一些特殊的课题时, 还必须亲自去搜集一些原始数据. 例如: 我们想了解消费者对某厂生产的凉风牌电扇的评价时, 必须进行一项调查. 厂方可以在每箱产品内放上一份用户意见调查表. 调查表的设计主要根据长期积累的工作经验, 但是有几项原则似乎是应当遵循的:

(1) 必须说明调查的目的, 表明本厂的诚意, 使被调查者愿意填写;

(2) 调查项目的说明要清楚. 例如要了解消费者的购买动机时, 可以采用如下的格式:

您之所以购买本厂生产的凉风牌电扇是由于(对下面一个或几个空格填写√表示回答):

- |                                  |                                |
|----------------------------------|--------------------------------|
| <input type="checkbox"/> 亲戚朋友的推荐 | <input type="checkbox"/> 电视广告  |
| <input type="checkbox"/> 造型美观    | <input type="checkbox"/> 价格便宜  |
| <input type="checkbox"/> 据说质量较好  | <input type="checkbox"/> 偶然的選擇 |

(3) 避免使用意义不明确的字句。例如提：“您购买的电扇已经使用了多少天？”这样的问题，就会使消费者感到迷惑，究竟是要把累计使用时间折算成天数呢？还是一天中只使用一个小时也算作一天？也许有些人会干脆把购买了多长时间作为答案。

(4) 必须避免可能造成偏见的调查。例如：如果调查卡上写上：感谢您购买本厂生产的凉风牌电扇。该产品曾荣获轻工业部颁发的优质产品奖，曾出口北美和东南亚，您是否感到本厂产品有如下特点：

- |                                 |                                 |
|---------------------------------|---------------------------------|
| <input type="checkbox"/> 造型比较美观 | <input type="checkbox"/> 风力比较柔和 |
| <input type="checkbox"/> 转速比较稳定 | <input type="checkbox"/> 噪音比较小  |

那么，这与其说是对消费者的意见的调查，还不如说是推销产品的广告。

调查表回收后，对各项回答进行统计就可以得到有用的信息。例如：在 1 000 份回收的调查表中若有 350 份，即 35% 回答说是由于看了电视广告而购买的话，那末说明广告宣传起了很大作用，应进一步利用这一手段。

统计资料中常常存在误差。我们且不谈在汇总过程中的计算误差，即使在“原始”的观测中，也会有种种误差发生。比如说，据统计 1983 年的刑事案件数目为 61 万件。无疑这指的是在公安机关登录在案的数目，不包括那些未报案或未被发现的案件，当然也不能否定存在错报的案件。以税收统计来说，几乎每个国家都存在偷税漏税的情况，当年的税收统计一年以后常常会被修正，原因之一是在那以后又查出了一些逃税事件。

原始数据经过按各种要求、特征的汇总处理后，就可以得到各种统计数据。例如：1982 年我国进行了第 3 次人口普查，数据的处理用了几十台电子计算机，历时 2 年半，最后整理成长达 600 多页的报告“中国 1982 年人口普查资料”。其中包括了大量关于人口特征，诸如年龄结构、性别结构、民族结构、区域结构、文化结构、职业结构以及家庭、婚姻、生育情况等的详细数据。比如说，到 1982 年

7月1日零时,我国总人口数为1 031 887 961人,其中,男性占51.33%,女性占48.67%,汉族占93.3%,少数民族占6.7%。

对某个统计指标在不同时期进行观测,将得到的数据按时间先后次序进行排列,这样得到的统计数据称为时间序列数据。每月的销售额、每季度的进口额、每年末的存款余额等都是时间序列数据。与此不同,若某个指标在不同的个体上进行观测,则得到该指标的一组横截面数据,上述的人口普查以及工业普查、住户调查等都可得到横截面数据。下面是典型的例子。

时间序列数据:全国钢产量

年 份	1980	1981	1982	1983	1984	1985	1986	1987	1988
产量(万吨)	3712	3560	3716	4002	4347	4679	5220	5628	5943

横截面数据:1988年华北各省市钢产量

地 区	北京	天津	河北	山西	内蒙古
产量(万吨)	369.0	156.4	308.7	216.1	220.8

## § 1.2 图形描述

### 一、统计资料的描述

对于统计资料,如上述历年钢产量或钢产量的地区分布等,除用表格形式描述外,还可以用各种图形来表现。一般说来,一张好的图形可以给人们带来很多信息。图形的好处是直观,可以启发人们的想像力,但它不一定能提供精确的数据。

在统计图形中,用得较多的是动态比较图和构成比较图。前者用于描述时间序列而后者较多地用于描述横截面数据。

图 1.1 是历年钢产量的动态线形图,它表明了钢产量的历年变化,从中可以对其增长的趋势以及变动大小产生一种直观的认识。



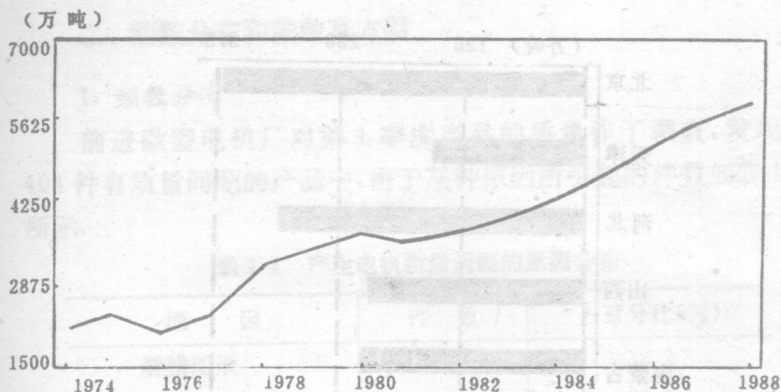


图 1.1 历年钢产量

识. 图 1.2 是一种条形动态图, 描绘的也是历年钢产量的变化, 但它给人的直观更强烈. 图 1.3 是一种水平条形图, 用以比较各地区的钢产量. 当要描述构成比时, 扇形图是最常用和方便的方法, 扇形的角度乘以  $100/360$  便是相应的构成比(图 1.4). 把历年的扇形图加以比较, 还可以了解某种结构的变动.

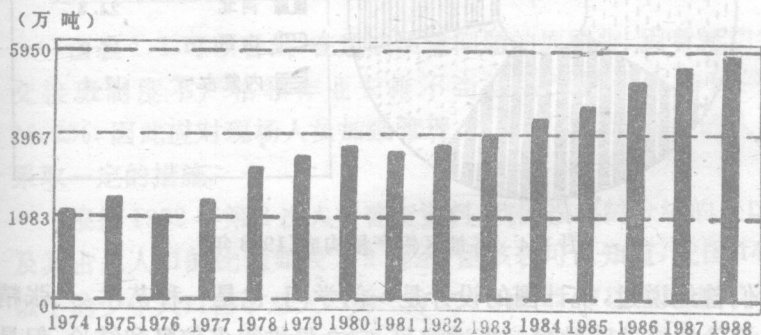


图 1.2 历年钢产量



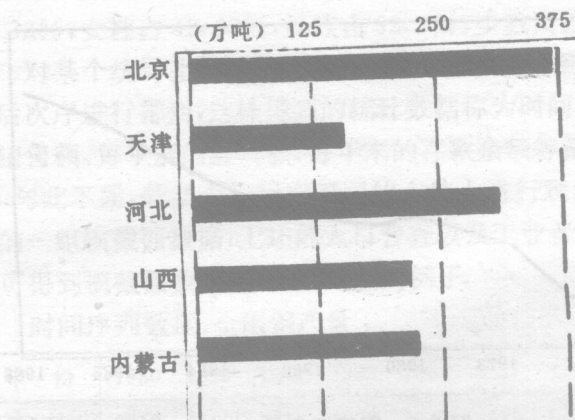


图 1.3 华北各地区钢产量(1988年)

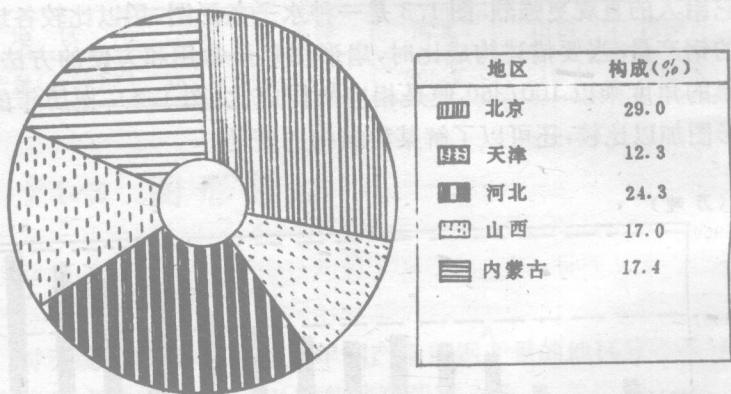


图 1.4 各地区钢产量构成(1988年)

我们曾经说过,统计图的设计是一门学问,也是一种艺术,一张精心设计的统计图胜过一篇讲演稿,它会提供很多信息和结论.但是在本书里,我们将不着重介绍各种描述统计图形的设计,我们想强调的是归纳统计方法.因此,我们在讲述描述统计时也只侧重于归纳统计有关的一些内容.

## 二、频数分布和频数直方图

### 1. 频数分布

前进微型电机厂对第1季度产品的质量作了调查,发现在404件有质量问题的产品中,由于某种原因所引起的件数如表1.1所示.

表 1.1 产生电机质量问题的原因分布

原 因	件 数	占百分比(%)
看错图纸	130	32.2
交接班不清楚	110	27.2
作业安排不当	70	17.3
设备功能问题	30	7.4
表面处理	18	4.5
喷涂	15	3.7
装配	6	1.5
其它	25	6.2
总 计	404	100.0

由表1.1可以看出,在造成质量问题的原因中,因看错图纸、交接班制度不严格和作业安排不当等人为错误占的比例达76.7%.因此应对现场人员加强管理,找出发生错误较多的人员,采取一定的措施.

根据1982年第3次人口普查资料,我国按年龄分组的人口数及其占总人口的比例如表1.2所示.由该表可以知道,我国14岁以下的少年人口占了总人口的三分之一,而65岁以上的老年人口则不到总人口的5%.同西方发达国家相比,据1981年的统计,14岁以下的人口比例,日本为23.4%,美国为22.3%,联邦德国为19.5%,法国为22.0%,英国为20.3%,都低于我国;相反,65岁以上的老年人口比例,日本为9.3%,美国为11.5%,联邦德国为

15.4%，法国为13.5%，英国为15.3%，都远远高于我国。因此，我国的人口是“年轻型”的，老龄化所产生的各种经济和社会问题在我国还不突出，而教育倒是应该引起特别重视的问题。

表 1.2 人口的年龄分组构成

年 龄 组	人口数(百万人)	占总人口比例(%)
0—14 岁	337.24	33.5
15—34 岁	369.18	36.6
35—49 岁	150.27	14.9
50—64 岁	102.18	10.1
65 岁以上	49.28	4.9
总 计	1008.15	100.0

资料来源：中国统计年鉴 1986。

上面的两张表我们称之为频数分布表。频数分布表可以为我们提供分析问题的重要信息，因此有着广泛的用途。这两张表的不同之点是：关于造成质量问题的原因的分类是定性的，而按年龄组的分类是定量的。在经济和管理问题中，有些事物的特征是可以量化的，例如：产值、劳动生产率、年龄、收入等；但是有些却无法定量，例如：文化程度、职业、产品的外观、纺织品的手感等。但是不管怎样，只要这些特征能按某种方式进行分类，就可以应用频数分布表。

## 2. 频数分布表的编制方法

由于定性特征的频数分布表的编制比较简单，这里我们只说明定量特征的频数分布表的编制方法。

表 1.3 是抽样调查 30 个工人家庭月收入的原始数据，这些数据可以记为  $x_1, x_2, \dots, x_{30}$ 。对于这些观测数据，

(1) 确定最大值  $x^{**}$  和最小值  $x^*$ 。根据表 1.3，月收入最高的是编号为 3 的家庭，最低的是编号为 29 的家庭，因此

$$x^{**} = 556, \quad x^* = 330.$$