

 高等学校现代统计学系列教材

工业 统计学

张崇岐 燕 飞

Industrial Statistics



高等教育出版社
HIGHER EDUCATION PRESS

高等学校现代统计学系列教材

Industrial Statistics

工业统计学

Gongye Tongjixue

张崇岐 燕 飞



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容提要

本书主要是为理工科大学“工业统计学”课程编写的教材,全书以统计学为主干内容,穿插一些概率论的基础知识,每章配以适当的练习题,内容由浅入深,便于学生更好地理解统计学的内容和加深对统计学本质的理解。

本书内容丰富,选材适当,便于教学,可作为高等学校理工科各专业的教材,也可供工程技术人员参考阅读。

图书在版编目(CIP)数据

工业统计学/张崇岐,燕飞编. —北京:高等教育出版社,2013.12

ISBN 978-7-04-038895-4

I. ①工… II. ①张… ②燕… III. ①工业统计学-高等学校-教材 IV. ①F402.4

中国版本图书馆 CIP 数据核字(2013)第 289257 号

策划编辑 张晓丽
插图绘制 尹莉

责任编辑 张晓丽
责任校对 窦丽娜

封面设计 赵阳
责任印制 刘思涵

版式设计 余杨

出版发行 高等教育出版社
社址 北京市西城区德外大街4号
邮政编码 100120
印刷 唐山市润丰印务有限公司
开本 787mm×960mm 1/16
印张 17.5
字数 320千字
购书热线 010-58581118

咨询电话 400-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版次 2013年12月第1版
印次 2013年12月第1次印刷
定价 27.60元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物料号 38895-00

高等学校现代统计学系列教材编委会

(按姓氏笔画排序)

主 编：方开泰

副主编：史宁中 何书元 陈 敏 耿 直

编 委：马 洪 方开泰 史宁中 杨 虎 何书元 何晓群
张爱军 张崇岐 陈 敏 郑 明 赵彦云 耿 直
曾五一 缪柏其

总 序

统计学是一门收集、整理和分析数据的科学和艺术。这里的“数据”通指“信息的载体”，涵盖了大千世界中的文本、图像、视频、时空数据、基因数据等。统计学是一个独立的学科，在历史上曾隶属于数学，但统计学与数学有着本质的区别，因此统计学教育有其自身的特点和要求，这些特点表现为：(1) 统计学研究的是随机现象，而数学研究的是确定性的规律；(2) 统计学是一门应用性很强的学科，许多概念和原理来自于实际的需要，不是数理逻辑的产物；(3) 数据在统计学中扮演了重要的角色。目前，统计学已被列为一级学科。

在过去的 30 年中，随着生命科学、信息科学、物质科学、资源环境、认知科学、工程技术、经济金融和人文科学等众多学科的发展，产生了许多新的统计学分支，如风险管理、数据挖掘、基因芯片分析等。此外，计算机及其有关软件在统计教育和应用中扮演了越来越重要的角色，它们提供了越来越多的图形表达和分析的方法，使得许多原来教科书中重要的内容，现在已变得无足轻重。统计教育必须要改革才能适应高速发展的形势。

大学的统计教育可分为两大类，一类是非统计学专业的课程，另一类是统计学专业的教学设计。非统计学专业的学生学习统计的目的是为了应用，在大学阶段，课程不多，主要是学习基础的统计概念和方法，学会使用统计软件，培养其解决实际问题的能力。统计学专业的课程设置十分重要，应向国际靠拢，对教师队伍的要求也较高。虽然这两类学生的教育有很多共同点，但在课程设置中必须加以区分。

我国的统计教育在过去受苏联的影响很深，把统计学作为数学的一个分支，在内容上偏理论，少应用，过于强调概率论在统计中的作用。统计学是一门应用性很强的学科，应从实际问题、从数据出发，通过统计的工具来揭示数据内部的规律。用“建模”的思路来教统计，使学生能更加容易理解统计的概念和方法，知道如何将实际问题抽象为统计模型，反过来又指导实践。对非统计学专业的学生，要强调统计的应用。学生要能熟练地使用至少一个统计软件包。对于统计学专业的学生，要培养学生对实际问题的建模能力。有些实际问题可直接应用现有的统计方法来解决，如问卷调查的统计分析。有些问题在初次接触时并不像一个统计问题，必须有坚实的

统计基础和对实际问题的洞察力,才能从中发掘出统计模型。要培养学生的这种能力及统计思想(统计思想是统计文化的一部分,是用统计学的逻辑思考问题)。教师在授课中要结合较多的应用例子,要求学生做案例研究,鼓励学生参加建模比赛,参加企业的实际项目。

为满足我国统计教育发展的需要,我们计划编写一套面向高校本科生、特别是一般院校,适用于统计学专业和非统计学专业的系列教材。系列教材的编写宗旨是:突出教学内容的现代化,重视统计思想的介绍,适应现代统计教育的特点及时代发展的新要求;以统计软件为支撑,注重统计知识的应用;内容简明扼要,生动活泼,通俗易懂。编写原则为:(1)从数据出发,不是从假设、定理出发;(2)从归纳出发,不是从演绎出发;(3)强调案例分析;(4)重统计思想的阐述,弱化数学证明的推导。系列教材分为两个方向,一个面对统计学专业,另一个面对非统计学专业和应用统计工作者。

高等学校现代统计学系列教材是适应形势的要求,由高等教育出版社邀请专家组成“高等学校现代统计学系列教材编委会”负责选题、审稿,由高等教育出版社出版。

以上是我们编写这套教材的背景和理念,希望得到读者的支持,特别是高校领导和教学一线的教师的支持。我们希望使用这套教材的师生和读者多提宝贵意见,使教材不断完善。

高等学校现代统计学系列教材编委会

前 言

本书是为高等理工科学校编写的工业统计学课程教材。理工科学校培养的工程技术人员在设计开发新产品、制造并改善工程系统等工程实践中起着举足轻重的作用,而他们的理论指导就来源于统计学。统计学是一门收集、整理和分析数据的科学和艺术。目前统计分析方法已经在经济学、生物学、社会学等学科领域有着非常广泛的研究。

本书将统计学的概念与方法应用于工业领域,旨在为解决工业中的各种问题提供有力的工具。全书不仅较为全面系统地介绍了统计的基本原理、基本方法及其应用,而且例题和习题大多来自现实生活,方便学生理解和掌握相关知识,有利于提高学生对统计学方法的应用能力。阅读本书需要一定的高等数学基础知识,希望读者在对工业统计方法理解的基础上学会如何应用。

全书共分 13 章,前两章简单介绍了统计学的主要研究内容以及统计学在工程中的作用,包括建立实证模型、设计工程试验、监控生产过程等。第 3 章介绍数据的表示方法,包括茎叶图、直方图、箱形图、时间序列图、概率图等,这些常用的数据表示方法在工业中经常遇到,理工科学生要熟悉并掌握。第 4 至 7 章主要介绍了概率、离散型和连续型随机变量的分布以及联合概率分布的有关知识。这四章的内容均为概率论的基础知识,尤其在 5,6 两章,介绍了工程实际遇到的不同概率分布,较为详细地讲述了这些概率分布的由来以及彼此之间的联系,更加方便读者理解记忆。第 8 至 13 章介绍统计学的相关概念和知识,尤其是统计推断部分。第 8 章简单介绍了工业生产中常用的统计量及其抽样分布,着重介绍样本均值的抽样分布。第 9 章先引入参数点估计的相关概念、重要性质以及比较方法,再介绍常见点估计方法:矩法估计和最大似然估计,最后一节简单介绍了基于贝叶斯统计学派的贝叶斯估计方法。第 10 章将点估计扩展到单个总体参数的统计区间,详细介绍总体均值、总体方差、总体比例的置信区间以及未来观测值的预测区间和正态分布的容许区间。第 11 章介绍单个总体参数的假设检验。第 12 章扩展到对两个总体的置信区间和假设检验的讨论。第 13 章介绍一元线性回归,包括最小二乘估计的方法与性质、回归方程的显著性检验以及回归预测问题。书后附录中简单

介绍了 MATLAB 在统计中的应用,有兴趣的读者可以将这一部分作为选学内容。

本书中涉及的部分数学知识只给出结论,省略了烦琐的定理证明和理论推导。但是这些证明在现有的教科书中都能找到,有兴趣深入了解的读者可以参阅书后列出的文献。

由于编者水平有限,书中难免有错漏以及欠妥之处,热诚地希望使用本书的教师和同学指正,以便再版时修改。

张崇岐 广州大学数学与信息科学学院, cqzhang@gzhu.edu.cn

燕 飞 广州大学数学与信息科学学院

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》,其行为人将承担相应的民事责任和行政责任;构成犯罪的,将被依法追究刑事责任。为了维护市场秩序,保护读者的合法权益,避免读者误用盗版书造成不良后果,我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为,希望及时举报,本社将奖励举报有功人员。

反盗版举报电话 (010)58581897 58582371 58581879

反盗版举报传真 (010)82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

目 录

第 1 章	统计学导论	1
1.1	什么是统计学	1
1.2	统计学的应用	2
	习题 1	3
第 2 章	统计在工程中的作用	5
2.1	工程方法和统计思想	5
2.2	收集工程数据	8
2.3	机械与实证模型	12
2.4	概率和概率模型	15
	习题 2	16
第 3 章	数据的表示方法	17
3.1	茎叶图	17
3.2	频率分布和直方图	19
3.3	箱形图	21
3.4	时间序列图	22
3.5	概率图	23
	习题 3	24
第 4 章	概率	26
4.1	样本空间和事件	26
4.1.1	随机试验	26
4.1.2	样本空间	27
4.1.3	事件	29
4.2	概率的定义	31
4.2.1	介绍	31
4.2.2	概率公理化定义	33
4.3	加法法则	33
4.4	条件概率	36
4.5	乘法法则和全概率法则	39
4.5.1	乘法法则	39

4.5.2 全概率法则	39
4.6 独立性	41
4.7 贝叶斯定理	44
4.8 随机变量	45
习题 4	46
第 5 章 离散随机变量及其概率分布	49
5.1 离散随机变量	49
5.2 概率分布列	50
5.3 累积分布函数	51
5.4 离散随机变量的均值和方差	53
5.5 离散均匀分布	56
5.6 二项分布	58
5.7 几何分布和负二项分布	63
5.7.1 几何分布	63
5.7.2 负二项分布	64
5.8 超几何分布	66
5.9 泊松分布	70
习题 5	73
第 6 章 连续随机变量及其概率分布	75
6.1 连续随机变量	75
6.2 概率密度函数	75
6.3 累积分布函数	78
6.4 连续随机变量的均值和方差	80
6.5 连续均匀分布	81
6.6 正态分布	83
6.7 二项分布和泊松分布的正态近似	88
6.8 指数分布	91
6.9 埃尔朗分布和伽马分布	95
6.9.1 埃尔朗分布	95
6.9.2 伽马分布	97
6.10 韦布尔分布	98
6.11 对数正态分布	99
习题 6	101

第 7 章	多维随机变量及其分布	104
7.1	二维离散随机变量	104
7.1.1	联合分布列	104
7.1.2	边际分布列	105
7.1.3	条件分布列	107
7.1.4	独立性	109
7.2	多维离散随机变量	111
7.2.1	联合概率分布	111
7.2.2	多项分布	113
7.3	二维连续随机变量	114
7.3.1	联合密度函数	114
7.3.2	边际密度函数	116
7.3.3	条件概率分布	118
7.3.4	独立性	120
7.4	多维连续随机变量	121
7.5	协方差和相关系数	124
7.6	二元正态分布	130
7.7	随机变量的线性组合	131
	习题 7	133
第 8 章	统计量及其抽样分布	136
8.1	总体和样本	136
8.2	统计量及其抽样分布	137
8.2.1	样本均值及其抽样分布	138
8.2.2	样本方差和样本标准差	141
	习题 8	143
第 9 章	参数点估计	145
9.1	点估计的一般概念	145
9.1.1	无偏估计	146
9.1.2	点估计的方差	147
9.1.3	点估计的标准误差	148
9.1.4	估计的均方误差	149
9.2	点估计的方法	150
9.2.1	矩法估计	150
9.2.2	最大似然估计	151
9.2.3	贝叶斯估计	156
	习题 9	159

第 10 章 单个总体参数的区间估计	161
10.1 区间估计	161
10.2 方差已知时, 正态总体均值的置信区间	162
10.2.1 置信区间的形成及其基本性质	162
10.2.2 样本容量的选择	164
10.2.3 单边置信界限	165
10.2.4 构造置信区间的一般方法	165
10.2.5 μ 的大样本置信区间	166
10.3 方差未知时, 正态总体均值的置信区间	168
10.3.1 t 分布	168
10.3.2 μ 的置信区间	169
10.4 正态总体方差的置信区间	171
10.5 二项分布参数 p 的大样本置信区间	173
10.6 未来观测值的预测区间	175
10.7 正态分布的容许区间	176
习题 10	177
第 11 章 单个总体参数的假设检验	179
11.1 假设检验	179
11.1.1 统计假设	179
11.1.2 统计假设的检验	180
11.1.3 单边和双边假设	185
11.1.4 假设检验的一般步骤	187
11.2 方差已知时, 正态总体均值的检验	187
11.2.1 均值的假设检验	187
11.2.2 假设检验中的 p 值	189
11.2.3 假设检验与置信区间的联系	190
11.2.4 第二类错误和样本容量的选择	190
11.3 方差未知时, 正态总体均值的检验	192
11.3.1 均值的假设检验	192
11.3.2 t 检验中的 p 值	194
11.4 正态总体方差的假设检验	194
11.5 总体比例的检验	196
11.5.1 总体比例的大样本检验	196
11.5.2 第二类错误和样本容量的选择	197

11.6 拟合优度检验	198
习题 11	202
第 12 章 两个总体的统计推断	204
12.1 方差已知时, 两个正态总体均值之差的推断	204
12.1.1 方差已知时, 两均值之差的假设检验	204
12.1.2 样本容量的选择	206
12.1.3 方差已知时, 两均值之差的置信区间	206
12.2 方差未知时, 两个正态总体均值之差的推断	207
12.2.1 方差未知时, 两均值之差的假设检验	207
12.2.2 方差未知时, 两均值之差的置信区间	212
12.3 配对 t 检验	213
12.4 两个正态总体方差的推断	216
12.4.1 F 分布	216
12.4.2 两方差比的假设检验	217
12.4.3 两方差比的置信区间	218
12.5 两个正态总体比率的推断	219
12.5.1 $H_0: p_1 = p_2$ 的大样本检验	219
12.5.2 $p_1 - p_2$ 的置信区间	221
习题 12	222
第 13 章 一元线性回归	224
13.1 一元线性回归模型	224
13.2 最小二乘估计及其性质	226
13.2.1 参数的最小二乘估计	226
13.2.2 最小二乘估计的性质	228
13.3 回归方程的显著性检验	230
13.3.1 t 检验	231
13.3.2 F 检验	231
13.4 区间估计和预测	233
13.4.1 回归系数的置信区间	233
13.4.2 观测值的预测区间	233
13.5 可化为线性回归的例子	234
习题 13	235
附录 MATLAB 在统计中的应用	237
附表 1 泊松分布数值表	251

附表 2	标准正态分布函数数值表	253
附表 3	t 分布分位数 $t_{\alpha,k}$ 表	254
附表 4	χ^2 分布分位数 $\chi_{\alpha,k}^2$ 表	255
附表 5	正态分布容许区间的系数 k 值表	256
附表 6	F 分布临界值表	257
参考文献		262

1.1 什么是统计学

“统计”一词,经常让我们联想到表格中的大量数字,一般情况下会认为“统计学”等同于“数据”。然而以数据作为统计学的概念解释并不符合其本身作为一门科学的意义,也不能对当代统计研究人员的活动进行确切地描述。下面我们将给出统计学的具体含义。

定义 1.1.1 统计学是一门收集、分析并解释数据的科学。

统计学的研究对象是大量偶然现象的数据特征和规律。例如某食品公司推出两种口味的糖果,从购买这两种糖果的消费者中随机抽取一部分人进行抽样调查,比较哪种口味的糖果更受欢迎;又如为了检查流水线上生产出的零件是否符合要求,要根据随机抽取出来的一定数量的零件进行推断。

统计学是一门年轻的科学,从 17 世纪中叶开始才逐渐形成系统的理论和方法。在统计学的不断发展中,由于在不同领域、不同行业所应用的统计方法不同,以及研究对象的差别等原因,统计学又衍生出了若干个分支学科和交叉学科。一般情况下,我们根据研究的侧重点不同,对统计学进行粗略的两种划分:一是分为描述统计学和推断统计学;一是分为理论统计学和应用统计学。

1. 描述统计学和推断统计学

描述统计学主要研究如何取得反映客观现象的数据,并通过图表形式对所收集的数据进行加工处理和显示,进而通过综合概括与分析得出反映客观现象的规律性数量特征。它包括统计数据的收集方法,数据的加工处理方法,数据的显示方法,数据分布特征的概括与分析方法等。

推断统计学主要研究如何根据样本数据去推断总体数量特征,它是在对样本数据进行描述的基础上,对统计总体的未知数量特征做出以概率形式表述的推断。它包括对部分数据的抽取,对数据的分析,根据部分数据对全体数据做出推断以及推断有效性的判断等内容。

统计研究过程的起点是统计数据,终点是探索出客观现象内在的数量规律性。在这一过程中,如果收集到的是总体数据(如普查数据),则经过描述统计

之后就可以达到认识总体数量规律性的目的了;如果所获得的只是研究总体的一部分数据(样本数据),要找到总体的数量规律性,则必须应用概率论的理论并根据样本信息对总体进行科学的推断.描述统计学和推断统计学的划分,一方面反映了统计方法发展的前后两个阶段,同时也反映了应用统计方法探索客观事物数量规律性的不同过程.描述统计是整个统计学的基础,推断统计则是现代统计学的主要内容.由于在对现实问题的研究中,所获得的数据主要是样本数据,因此,推断统计在现代统计学中的地位和作用越来越重要,已成为统计学的核心内容.当然,这并不等于说描述统计不重要,如果没有描述统计收集可靠的统计数据并提供有效的样本信息,即使再科学的统计推断方法也难以得出切合实际的结论.从描述统计学发展到推断统计学,既反映了统计学发展的巨大成就,也是统计学发展成熟的重要标志.

2. 理论统计学和应用统计学

理论统计学主要是把研究对象一般化、抽象化,以概率论为基础,从纯理论的角度,对统计方法加以推导论证,中心的内容是统计推断问题,实质是以归纳方法研究随机变量的一般规律.它包括概率论、抽样理论、样本分布理论、试验设计、估计理论、假设检验、决策理论、非参数统计、序贯分析、多元分析、时间序列以及博弈理论等.理论统计学着重阐明统计学的数学原理,为统计方法提供理论基础.

应用统计学是将统计方法应用于解决自然科学和社会科学领域中的实际问题时所产生的统计学.它不需要研究数学原理的推导,也不研究公式和定理的证明.它需要的就是基本的统计方法和所研究专业的知识,两者的结合就产生了各专业的统计学,如社会统计学、人口统计学、生物统计学、心理统计学、经济统计学、工业统计学等.

应用统计学和理论统计学有三点不同之处:第一,从研究对象来看,应用统计学是以客观现象中具体的指标数值作为主要研究对象;第二,从研究方法来看,应用统计学是研究如何运用统计学的理论和方法来研究现实问题;第三,从研究目的来看,应用统计学以了解掌握客观现象的数量特征、数量关系和统计规律为目的.

1.2 统计学的应用

统计学的应用非常普遍,几乎在社会生活的各个领域都能看见.它的影响也是巨大的,以至于无论怎样强调统计的重要性都不为过.

在政府中,统计资料被用来研究和影响政府在税收、公共设施建设(例如道