

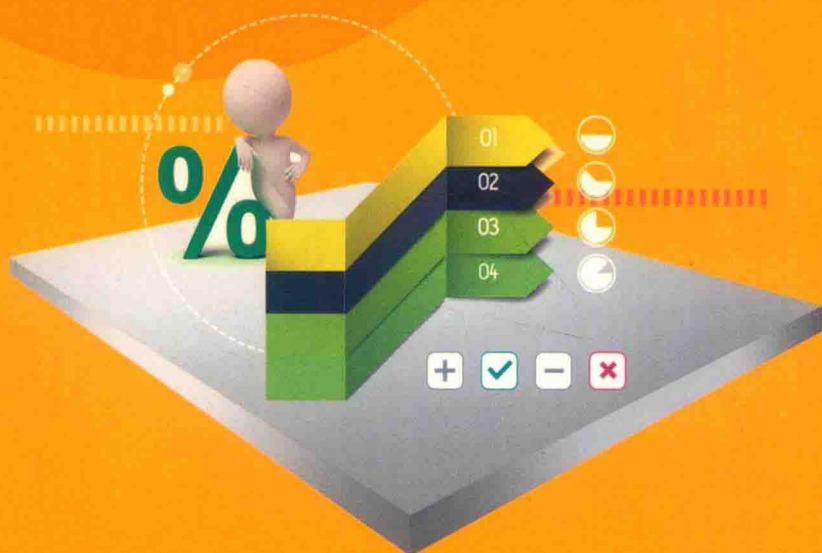
经济与管理专业精品教材

应用统计学

Y

主编 陈钦 王灿雄

INGYONG TONGJI XUE



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

经济与管理专业精品教材

应用统计学

主编 陈 钦 王灿雄



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书系统地阐述了应用统计学的相关知识,全书分为十章,主要包括绪论、统计调查、图表描述、描述性统计量、参数估计、假设检验、列联分析、方差分析、回归分析和时间序列分析等内容。

本书内容精练、体例活泼、实例丰富、突出实用,既可作为普通高等院校经济与管理专业及其他相关专业的教学用书,也可作为统计工作者的参考读物。

图书在版编目(CIP)数据

应用统计学 / 陈钦, 王灿雄主编. -- 上海: 上海交通大学出版社, 2017

ISBN 978-7-313-17830-5

I. ①应… II. ①陈… ②王… III. ①应用统计学—教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2017)第 186347 号

应用统计学

主 编: 陈 钦 王灿雄

出版发行: 上海交通大学出版社

邮政编码: 200030

出 版 人: 郑益慧

印 制: 北京市科星印刷有限责任公司

开 本: 787mm×1092mm 1/16

版 次: 2017 年 8 月第 1 版

书 号: ISBN 978-7-313-17830-5/C

定 价: 48.00 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 16.5 字 数: 271 千字

印 次: 2017 年 8 月第 1 次印刷

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与发行部联系

联系电话: 010-62137141



前言 | PREFACE

在经济全球化和信息化的环境下,管理决策者面临着高度不确定的外部环境和巨大的风险。应用统计学是定量分析非确定性问题的规律,帮助管理决策者进行科学决策、规避风险、获得最优经济效益和社会效益的科学方法,已成为现代科学管理中必不可少的强有力的工具。

本书系统地阐述了应用统计学的相关知识,全书分为十章,主要包括绪论、统计调查、图表描述、描述性统计量、参数估计、假设检验、列联分析、方差分析、回归分析和时间序列分析等内容。

与其他同类教材相比,本书具有以下特色:

- **内容精练:** 本书在确保知识的系统性和准确性的基础上,尽量使用通俗易懂的语言,阐述应用统计学的基本概念、基本原理、解决问题的基本方法及应用条件,略去了繁琐的推导证明过程。
- **体例活泼:** 本书力求突破传统教材的说教形式,每章内容都以“引导案例”引出本章重点知识;在知识讲解中插入了“小提示”“延伸阅读”“课堂讨论”等模块,既可为学生营造出更加直观、立体的认知环境,又可引导学生进行自主学习;每章都设置有“思考与练习”,便于学生复习、巩固所学知识,检验学习成效。
- **实例丰富:** 本书融入了大量操作性较强的实例,并对实例进行了详细分析,使学生更容易理解和掌握理论知识。
- **突出实用:** 鉴于计算机是求解统计问题的主要工具,本书对主要的统计方法,都以常用的统计软件 Excel 或 SPSS 作为求解运算和分析的主要工具,既可提高本书的实用性,又有助于提高学生的计算机应用能力。

本书是福建农林大学金山学院第一批应用型课程建设项目“统计学”(项目编号:K170208)的成果。本书由陈钦和王灿雄主编,具体分工如下:陈钦编写第一章、第二章及第五至第八章;王灿雄编写第三章、第四章、第九章、第十章及附录。在编写过程中,我们参考了大量的文献资料。在此,我们向参考过的文献的作者表示诚挚的谢意。

由于编写时间仓促,编者水平有限,书中疏漏与不当之处在所难免,敬请广大读者批评指正。

编 者

2017年8月

» 目 录

第一章 绪论	1
引导案例	2
第一节 统计学概述	2
一、统计与统计学的概念	2
二、统计学的研究方法	3
三、统计学与其他学科的关系	5
第二节 统计数据类型	6
一、按所采用的计量尺度划分	6
二、按数据收集方法划分	7
三、按描述对象与时间的关系划分	8
第三节 统计学的基本概念	8
一、总体和样本	8
二、参数和统计量	10
三、标志和指标	11
四、变量	13
本章小结	14
思考与练习	15
第二章 统计调查	17
引导案例	18
第一节 统计调查概述	19
一、统计调查的概念	19
二、统计调查的意义	20

第二节 统计调查的组织方式	20
一、普查	20
二、重点调查	22
三、典型调查	23
四、抽样调查	24
第三节 统计调查的方法	27
一、直接观察法	27
二、报告法	27
三、采访法	28
四、通讯法	28
五、实验调查法	29
第四节 统计调查问卷	30
一、调查问卷的类型	30
二、调查问卷的内容	31
本章小结	34
思考与练习	35
第三章 图表描述	37
引导案例	38
第一节 品质型数据的图表描述	39
一、单项式频数分布表	39
二、条形图与饼形图	40
第二节 数值型数据的图表描述	43
一、组距式频数分布表	43
二、直方图	46
三、盒形图	47
四、茎叶图	48
五、频数分布的类型	49
第三节 双变量关系的图表描述	50
一、交叉频数分布图	51
二、散点图	51
第四节 运用 SPSS 进行图表描述	53
一、运用 SPSS 制作简单频数分布表	53
二、运用 SPSS 制作条形图	54
三、运用 SPSS 制作饼形图	55

四、运用 SPSS 制作直方图	56
五、运用 SPSS 制作盒形图	56
六、运用 SPSS 制作茎叶图	57
七、运用 SPSS 制作交叉频数分布表和交叉频数分布图	58
八、运用 SPSS 制作散点图	59
本章小结	61
思考与练习	61
第四章 描述性统计量	65
引导案例	66
第一节 集中趋势描述性统计量	67
一、均值	67
二、中位数	70
三、众数	71
四、均值、中位数和众数的比较	71
第二节 离散程度描述性统计量	72
一、极差、四分位差与平均差	72
二、方差与标准差	75
三、离散系数	79
第三节 分布形态描述性统计量	80
一、偏度	80
二、峰度	80
第四节 运用 SPSS 进行统计量描述	81
一、由“Frequencies”计算描述统计量	82
二、由“Descriptives”计算描述统计量	83
三、由“Explore”计算描述性统计量	84
本章小结	85
思考与练习	86
第五章 参数估计	91
引导案例	92
第一节 参数估计的基本原理	92
一、估计量与估计值	93
二、点估计与区间估计	93
三、置信区间与置信水平	94
四、几种重要的随机变量分布	94

第二节 一个总体参数的区间估计	97
一、总体均值的区间估计	97
二、总体比例的区间估计	100
三、总体方差的区间估计	101
第三节 两个总体参数的区间估计	102
一、两个总体均值之差的区间估计	102
二、两个总体比例之差的区间估计	108
三、两个总体方差之比的区间估计	109
第四节 样本量的确定	110
一、估计总体均值时样本量的确定	110
二、估计总体比例时样本量的确定	111
本章小结	112
思考与练习	112
第六章 假设检验	117
引导案例	118
第一节 假设检验的基本问题	118
一、假设检验的原理	118
二、假设检验的流程	119
第二节 一个总体参数的检验	122
一、总体均值的检验	122
二、总体比例的检验	125
三、总体方差的检验	126
第三节 两个总体参数的检验	127
一、两个总体均值之差的检验	127
二、两个总体比例之差的检验	134
三、两个总体方差之比的检验	135
本章小结	136
思考与练习	137
第七章 列联分析	139
引导案例	140
第一节 列联表	141
一、列联表的构造	141
二、列联表的分布	141

第二节 χ^2 统计量与 χ^2 检验	143
一、 χ^2 统计量	143
二、 χ^2 检验	144
第三节 列联表中的相关测量	146
一、 ϕ 相关系数	146
二、 c 相关系数	147
三、 V 相关系数	147
本章小结	148
思考与练习	149
第八章 方差分析	151
引导案例	152
第一节 方差分析概述	152
一、方差分析中的相关术语	153
二、方差分析的基本原理	154
第二节 单因素方差分析	155
一、数据结构	155
二、分析步骤	155
三、方差分析表	159
四、用 Excel 工具进行方差分析	159
五、关系强度的测量	160
六、方差分析中的多重比较	161
第三节 双因素方差分析	163
一、双因素方差分析的类型	163
二、无交互作用的双因素方差分析	163
三、有交互作用的双因素方差分析	167
本章小结	170
思考与练习	170
第九章 回归分析	173
引导案例	174
第一节 相关分析概述	175
一、相关关系	175
二、相关系数	177
三、相关系数的显著性检验	180

第二节 回归模型与回归方程	181
一、回归模型	182
二、回归方程	182
第三节 估计的回归方程	184
一、估计回归方程的概念	184
二、参数的最小二乘估计	184
三、判定系数	186
第四节 检验和预测	188
一、 F 检验	188
二、回归预测	190
第五节 残差分析	196
一、残差图	196
二、标准化残差图	198
三、正态概率图	200
四、异常值的检测	201
第六节 运用 SPSS 进行回归分析	203
一、运用 SPSS 进行一元线性回归分析	203
二、曲线估计	207
本章小结	210
思考与练习	210
第十章 时间序列分析	215
引导案例	216
第一节 时间序列的构成	217
一、长期趋势	217
二、季节波动	218
三、循环波动	218
四、无规则波动	218
五、平稳序列与非平稳序列	219
第二节 波动性的描述	219
一、逐期增减量与累积增减量	219
二、环比发展速度与定基发展速度	220
三、环比增减速度与定基增减速度	221
第三节 长期趋势的测定	222
一、平均移动法	222



二、最小平方法	225
第四节 季节波动的测定	226
一、同季平均法	226
二、趋势剔除法	228
三、季节调整	230
本章小结	231
思考与练习	232
附录 用 Excel 生成概率分布表	236
附录一 标准正态分布表	236
附录二 标准正态分布分位数表	238
附录三 t 分布临界值表	240
附录四 χ^2 分布临界值表	242
附录五 F 分布临界值表 ($\alpha=0.05$)	244
附录六 r 简单相关系数临界值表	246
附录七 随机数字表	248
参考文献	250



第一章 绪 论

内容提要

统计学是关于数据的学科，它所提供的是一套数据的收集、处理、分析、解释，并从数据中得出结论的方法。统计学的力量在于无处不在的应用性，无论人们从事什么工作，都可能遇到如何收集有价值的资料，如何组织、解释所收集的资料，如何分析并给出适当的推论以及推论的可信度有多高等问题，这些都是统计学的应用范畴。

学习目标

知识目标

- 熟悉统计学的研究方法
- 掌握统计数据的类型
- 掌握统计学的基本概念

能力目标

- 能区分统计数据的不同类型
- 能区分总体和样本、参数和统计量、标志和指标、变量的概念

引导案例

大数据时代带来的大变革

国际数据公司定义了大数据的特征：海量的数据规模、快速的数据流转、动态的数据体系、多样的数据类型和巨大的数据价值。仅从海量的数据规模来看，全球 IP 流量达到 1 EB 所需的时间，在 2001 年需要 1 年，在 2013 年仅需 1 天，到 2016 年则仅需半天。全球新产生的数据年增 40%，全球信息总量每两年就可翻番。

而根据 2012 年互联网络数据中心发布的《数字宇宙 2020》报告，2011 年全球数据总量已达到 1.87 ZB (1 ZB=10 万亿亿字节)，如果把这些数据刻成 DVD，排起来的长度相当于从地球到月亮一个来回的距离，并且数据以每两年翻一番的速度飞快增长。预计到 2020 年，全球数据总量将达到 35~40 ZB。

需要强调的是，所谓大数据并不仅仅是指海量数据，而更多的是指这些数据都是非结构化的、残缺的、无法用传统方法进行处理的数据。基于互联网技术而发展起来的大数据应用，将会对人们的生产过程和商品交换过程产生颠覆性的影响。大数据的本质是基于互联网基础的信息化应用，其真正的“魔力”在于信息化与工业化的融合，使工业生产效率得到大规模提升。

肯尼思·库克耶和维克托·迈尔·舍恩伯格指出，数据出现了三大变化：第一，人们处理的数据从样本数据变成全部数据；第二，由于是全样本数据，人们不得不接受数据的混杂性，而放弃对精确性的追求；第三，人们通过对大数据的处理，放弃对因果关系的渴求，转而关注相互联系。这一切代表着人类告别总是试图了解世界运转方式背后深层原因的态度，而走向仅仅需要弄清现象之间的联系并利用这些信息来解决问题。我们急需拥有从各种数据中快速获得有价值的信息的能力，统计学就是培养这种能力的学科。

第一节

统计学概述

一、统计与统计学的概念

统计作为一种社会实践活动，已有悠久的历史。可以说，自从有了国家，就有了统计



实践活动。最初,统计只是一种计数活动,为满足统治者管理国家的需要而收集资料,通过统计计数弄清国家的人力、物力和财力,作为国家管理的依据。现在,“统计”一词已被人们赋予多种含义,它可以指统计数据的收集活动,即统计工作;也可以指统计活动的结果,即统计数据;还可以指分析统计数据的方法和技术,即统计学。

统计学是指收集、处理、分析、解释数据,并从数据中得出结论的科学。统计学是关于数据的科学,统计研究的是来自各领域的数据。

二、统计学的研究方法

社会经济现象的复杂性决定了统计学研究方法的多样性,在统计研究过程的各个阶段,必须运用各种专门的方法。统计学研究方法主要包括大量观察法、实验设计法、统计分组法、综合指标法、指数分析法、抽样推断法和统计模型法等。

(一) 大量观察法

所谓大量观察法,是指对所研究的事物的全部或足够多的数量进行观察的方法。社会现象或自然现象都受到各种社会规律或自然规律的相互交错作用的影响。在现象总体中,个别单位往往受偶然因素的影响,如果任选其中之一进行观察,其结果不足以代表总体的一般特征。只有观察全部或足够多的单位并加以综合,影响个别单位的偶然因素才会相互抵消,现象的一般特征才能显示出来。

大量观察法的理论基础是大数定律,其作用主要在于通过对统计总体的大量单位进行观察,使大量个体中非本质的偶然因素的影响相互抵消或削弱,从而将统计总体的本质特征显示出来,达到正确认识客观事物发展规律的目的。例如,在投掷骰子的游戏中,随机投掷匀质的骰子,哪个面会出现是不确定的、完全偶然的。但只要进行多次重复投掷,就会发现每个面出现的次数大体相同,即比值接近 $1/6$ 。投掷的次数越多,就越接近 $1/6$ 这一稳定的数值。这里的 $1/6$ 就是投掷骰子出现某一特定结果的概率,也就是投掷骰子时所呈现的数量规律性。

(二) 实验设计法

实验设计法是指根据统计研究的目的,事先设计合理的实验程序,在实验过程中排除可控因素的差异,显示出不可控(随机)因素的差异,使得收集到的统计数据符合统计分析方法的要求,以便有效地揭示客观现象的特征。实验设计法主要适用于工程技术、生物医药、心理学、教育学、管理学等领域的统计数据的收集。进行实验设计时,要遵循以下三个原则。

1. 重复性原则

在相同条件下进行多次重复实验,既可以验证实验方案的可行性,揭露实验过程中的

不规范行为;又可以把各次实验的结果数据加以综合,更加逼近客观现象的数据特征(真值),减少误差,取得满意的实验结果。

2. 随机化原则

在实验过程中,实验对象单位的分配和次序都是随机安排的,这种安排可以使实验过程中的可控因素均匀化,尽量减少可控因素的影响,突出不可控因素的影响。

3. 匹配原则

实验设计的基本逻辑是:有意识地改变某个变量(设为A项)的情况,然后看另一个变量(设为B项)变化的情况。如果B项随着A项的变化而变化,就说明A项对B项有影响。为此,需要将研究对象分为两组,即实验组和对照组。实验组是指随机抽选的实验对象的子集,在这个子集中,每个单位接受某种特别的处理;而在对照组中,每个单位不接受实验组成员所接受的某种特别的处理。一个好的实验,对照组和实验组的产生不仅是随机的,而且应该是匹配的。所谓匹配,是指对实验单位的背景材料进行分析比较,将情况类似的每对单位分别随机地分配到实验组和对照组。

(三) 统计分组法

统计分组法是指根据统计研究的目的和统计总体的内在特点,按一定的统计标志将总体划分为性质不同的组或类的方法。统计总体的变异性是统计分组的前提条件,由于研究现象本身的复杂性、差异性及多层次性,需要对研究现象进行分组或分类研究,以期在同质的基础上探求不同组或类之间的差异性。

统计分组在整个统计活动过程中都占有重要地位。在统计调查阶段,可通过统计分组来收集不同的资料,使抽样调查的样本的代表性得以提高;在统计整理阶段,可通过统计分组使各种数据资料得到分门别类的加工处理和储存,并为编制分布数列奠定基础;在统计分析阶段,则可通过统计分组来划分现象的类型、研究总体的内在结构、比较不同组或类之间的差异(显著性检验)和分析不同变量之间的相关关系。

(四) 综合指标法

综合指标法是指运用各种统计指标来反映和研究客观总体现象的一般数量特征和数量关系的方法。统计研究现象的数量方面的特征是通过统计综合指标来反映的。通过综合指标的计算,可以显示出现象在具体时间和地点条件下的总量规模、相对水平、集中趋势、变异程度,并进一步从动态上研究现象的发展趋势和变化规律。如何最真实客观地记录、描述和反映所研究现象的数量特征和数量关系,是统计指标理论研究的一大课题。

(五) 指数分析法

指数分析法主要用于描述和分析事物数量方面的相对变化程度。由于现象的总体非常复杂,发展变动受其构成要素变动的影响,而这些构成要素往往具有不同的性质而不可以



直接相加,因此不能进行直接的比较观察。指数分析法正是通过同度量因素,把不能直接相加的要素转化为可以相加的现象总量,并计算出指数来反映复杂现象的总量的变动。另外,还可以对现象的各个构成要素进行因素分析,以显示它们的变动对总体变动的影响程度和方向。

(六) 抽样推断法

统计学在研究现象的总体数量特征时,需要了解的总体对象的范围往往是很大的,有时甚至是无限的。而由于经费、时间和精力等各方面的原因,以致在客观上难以实现,往往只能对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,要说明某种产品的平均使用寿命,只能从该产品中抽取一小部分进行检验,推断这种产品的平均使用寿命,并给出这种推断的置信水平。这种在一定置信水平下,根据样本资料的特征对总体特征作出估计和预测的方法称为抽样推断法。抽样推断法已在统计研究的许多领域得到应用,除了最常见的总体指标推断外,统计模型参数的估计和检验、统计预测中时间序列的估计和检验等也都属于统计推断的范畴。

(七) 统计模型法

在以统计指标反映所研究现象的数量特征的同时,还需要经常对相关现象之间的数量变动关系进行定量研究,以了解某一(些)现象数量变动与另一(些)现象数量变动之间的关系及变动的影响程度。统计模型法是综合指标法的扩展,是指根据一定的理论和假定条件,用数学方程模拟客观现象相互关系的一种研究方法。运用这种方法,可以对客观现象和过程中存在的数量关系进行比较完整和全面的描述,凸显所研究的综合指标之间的关系,从而简化客观存在的其他复杂关系,以便利用模型对所关心的现象变动进行评估和预测。

三、统计学与其他学科的关系

统计学是一门应用性很强的学科。由于几乎所有的学科都要研究和分析数据,因而统计学与这些学科领域都有着或多或少的联系。

(一) 统计学与数学的关系

统计学与数学有着密切的关系。由于现代统计学用到了较多数学知识,研究理论统计学的人需要有较深的数学功底,应用统计方法的人也要具备良好的数学基础,这就给人造成了一种错觉,似乎统计学是数学的一个分支,这种理解是不妥当的。实际上,数学只是为统计理论和统计方法的发展提供了数学基础,而统计学的主要特征是研究数据;另一方面,统计方法与数学方法一样,并不能独立地直接研究和探索客观现象的规律,而是给各学科提供了一种研究和探索客观规律的数量方法。

统计学与数学又有着本质的区别。首先,虽然从表面上看,统计学与数学都是研究数量规律的,但实际上二者却有着明显的区别。数学研究的是抽象的数量规律;而统计学则是研究具体的、实际现象的数量规律。其次,统计学与数学研究中所使用的逻辑方法不同。数学研究使用的是纯粹的演绎,而统计学则是演绎与归纳相结合,占主导地位的是归纳。数学家可以凭借聪明的大脑从假设命题出发推导出结果,而统计学家则需要深入实际收集数据,并与具体的实际问题相结合,经过科学的归纳才能得出有益的结论。

(二) 统计学与会计学的关系

统计学和会计学都是具有典型计量特征的应用科学,二者都是以数据为语言,揭示其研究对象的特征,并实行有效的监督或管理。但是,会计学主要应用于社会经济领域,而统计学的应用领域更为广阔。

社会经济的信息主要分为会计信息和统计信息两大类,会计学主要提供货币尺度方面或价值指标的信息,统计学则提供多种指标尺度方面的信息。但是,统计学中所处理的货币方面的信息,又是由会计学提供的;会计学中所用的实物量指标等,又是由统计学提供的。

由于统计对象的多样性和大量性,学科上也就有社会经济统计、生物统计、气象统计等分支,并且统计方法更多地应用数理方法。当统计对象不便或不需精确计量时,抽样、概率、模糊等数理方法得到广泛应用,一些数理统计方法也广为会计学所用。



课堂讨论

CLASS DISCUSSION

除数学与会计学外,还有哪些学科与统计学联系密切?

第二节

统计数据类型

统计数据是指对客观现象进行测量的结果。根据不同的标准,统计数据可分为不同的类型。

一、按所采用的计量尺度划分

按所采用的计量尺度划分,统计数据可分为分类数据、顺序数据和数值型数据。

(一) 分类数据

分类数据是指只能归于某一类别的非数字型数据,是对事物进行分类的结果。分类数