

大学计算机基础与应用系列立体化教材

# 统计数据分析基础教程

——基于SPSS 20 和Excel 2010的调查数据分析（第二版）

叶 向 李亚平◎编著



中国人民大学出版社

■ 大学计算机基础与应用系列立体化教材

# 统计数据分析基础教程

——基于SPSS 20 和Excel 2010的调查数据分析（第二版）

叶 向 李亚平◎编著



中国人民大学出版社

· 北京 ·

## 图书在版编目 (CIP) 数据

统计分析基础教程——基于 SPSS 20 和 Excel 2010 的调查数据分析/叶向, 李亚平编著. —2 版. —北京: 中国人民大学出版社, 2015. 4  
大学计算机基础与应用系列立体化教材  
ISBN 978-7-300-21089-6

I. ①统… II. ①叶… ②李… III. ①统计数据-统计分析-应用软件-高等学校-教材 IV. ①O212.1-39

中国版本图书馆 CIP 数据核字 (2015) 第 069426 号

大学计算机基础与应用系列立体化教材  
**统计分析基础教程 (第二版)**  
——基于 SPSS 20 和 Excel 2010 的调查数据分析  
叶向 李亚平 编著  
Tongji Shuju Fenxi Jichu Jiaocheng

---

出版发行	中国人民大学出版社	邮政编码	100080
社 址	北京中关村大街 31 号		
电 话	010-62511242 (总编室)		010-62511398 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	<a href="http://www.crup.com.cn">http://www.crup.com.cn</a>		
	<a href="http://www.ttrnet.com">http://www.ttrnet.com</a> (人大教研网)		
经 销	新华书店	版 次	2010 年 2 月第 1 版
印 刷	北京鑫丰华彩印有限公司		2015 年 4 月第 2 版
规 格	185 mm×260 mm 16 开本	印 次	2015 年 4 月第 1 次印刷
印 张	22.25 插页 1	印 定	价 46.00 元
字 数	491 000		

---

版权所有 侵权必究 印装差错 负责调换

## 内容简介

本教材介绍了从设计调查问卷开始直到完成调查报告为止的整个社会调查过程,包括问卷设计与数据收集、问卷数据的录入与清理、问卷数据基本统计分析(单变量的一维频率分析、双变量的交叉表分析、多选变量的一维频率分析和交叉表分析、描述统计分析)、假设检验、单因素方差分析、线性相关分析与线性回归分析等内容。作者还结合自己 10 多年学习、实践、讲授和研究统计方法及其应用的经验体会,介绍了许多专门的方法和技巧。

本教材的写作基础是安装于 Windows 7 操作系统上的中文版 SPSS 20 和中文版 Excel 2010。

本教材可作为各级各类高等院校本科生统计数据分析的入门教材,也可以作为 MBA 学生、研究生以及从事统计数据分析工作的人士的参考书。同时本教材也十分便于实际调研部门的人员和对数据分析感兴趣的读者自学及实践时参考。

## 总序

随着计算机与互联网应用的普及、信息技术的发展及中小学对信息技术基础课程的普遍开设,针对大学计算机基础与应用教育的方向和重点,我们认为应该研究新的教育与教学模式,使得计算机基础与应用课程摆脱传统的课堂上课+课后上机这种简单、低效的教学方式,逐步转向以实践性教学和互动式教学为手段,利用现代化的计算机实现辅助教学、管理与考核,同时提供包括教材、教辅、教案、习题、实验、网络资源在内的丰富的立体化教学资源 and 实时或在线答疑系统,使得学生乐于学习、易于学习、学有所成、学有所用,同时减轻教师备课、授课、布置作业与考核、阅卷的工作量,提高教学效率。这是我们建设这套“大学计算机基础与应用系列立体化教材”的初衷。

根据大学非计算机专业学生的社会需求和教育部对计算机基础与应用教育的指导意见,中国人民大学从2005年开始对计算机公共课进行大规模改革,包括增设课程、改革教学方式和考核方式、进行教材建设等多个方面的内容。在最新的《中国人民大学本科生计算机教学指导纲要(2008年版)》中,将与计算机教育有关的内容分为三个层次。第一层次为“计算机应用基础”课程,第二层次为“计算机应用类”课程(包含约10门课程),第三层次纳入专业基础课或专业课教学范畴,形成1+X+Y的计算机基础与应用教育格局。其中,第一层次的“计算机应用基础”课程和第二层次的“计算机应用类”课程,作为分类分层教学中的核心课程,走在教学改革的前列,同时结合中国人民大学计算机教学改革中开展的其他项目,已经形成了教材(部分课程)、教案、教学网站、教学系统、作业系统、考试系统、答疑系统等多层次、立体化的教学资源。同时,部分项目获得了学校、北京市、全国各级教学成果奖励和立项。

为了巩固我们的计算机基础与应用教学改革成果并使其进一步深化,我们认为有必要系统地建立一套更合理的教材,同时将前述各项立体化、多层次的教学资源整合到一起。为此,我们组织中国人民大学、中央财经大学、天津财经大学、河北大学、东华大学、华北电力大学等多所院校中从事计算机基础与应用课程教学的一线骨干教师,共同建设“大学计算机基础与应用系列立体化教材”项目。

本项目对中国人民大学及合作院校的计算机公共课教学改革和课程建设起着非常关键的作用,得到了各校领导和相关部门的大力支持。该项目将在原来的应用教学的基础上,更进一步地加强实践性教学、实验和考核环节,让学生真正地做到学以致用,与信息技术的发展同步成长。

本系列教材覆盖了“计算机应用基础”(第一层次)和“计算机应用类”(第二层次)的十余门课程,包括:

- 大学计算机应用基础

- Internet 应用教程
- 多媒体技术与应用
- 网站设计与开发
- 数据库技术与应用
- 管理信息系统
- Excel 在经济管理中的应用
- 统计数据分析基础教程
- 信息检索与应用
- C 程序设计教程
- 电子商务基础与应用

每门课程均编写了教材和配套的习题与实验指导。

随着信息化技术的发展，许多新的应用不断涌现，同时数字化的网络教学手段也在发展和成熟。我们将为此项目全面、系统地构建立体化的课程与教学资源体系，以方便学生学习、教师备课、师生交流。具体措施如下：

- 教材建设：在教材中减少纯概念性理论的内容，加强案例和实验指导的分量；增加关于最新的信息技术应用的内容并将其系统化，增加互联网和多媒体应用方面的内容；密切跟踪和反映信息技术的新应用，使学生学到的知识马上就可以使用，充分体现“应用”的特点。

- 教辅建设：针对教材内容，精心编制习题与实验指导。每门课程均安排大量针对性很强的实验，充分体现课程的实践性特点。

- 教学视频：针对主要教学要点，我们将逐步录制教学操作视频，使得学生的学习和复习更为方便。

- 电子教案：我们为教师提供电子教案，针对不同专业和不同的课时安排提出合理化的教学备课建议。

- 教学网站：纸质课本容量有限，更多更全面的教学内容可以从我们的教学网站上查阅。同时，新的知识、技巧和经验不断涌现，我们亦将它们及时地更新到教学网站上。

- 教学辅助系统：针对采用本教材的院校，我们开发了教学辅助系统。通过该系统，可以完成课程的教学、作业、实验、测试、答疑、考试等工作，极大地减轻教师的工作量，方便学生的学习和测试，同时网络的交流环境使师生交流答疑更为便利。（对本教学辅助系统有兴趣的院校，可联系 [yxd@yxd.cn](mailto:yxd@yxd.cn) 了解详情。）

- 自学自测系统：针对个人读者，可以通过我们提供的自学自测系统来了解自己学习的情况，调整学习进度和重点。

- 在线交流与答疑系统：及时为学生答疑解惑，全方位地为学生（读者）服务。

相信本套教材和教学管理系统不仅对参与编写的院校的计算机基础与应用教学改革起到促进作用，而且对全国其他高校的计算机教学工作也具有参考和借鉴意义。

杨小平

2009年6月

## 第二版前言

自从周以真教授比较系统地提出计算思维的概念以来,计算思维便受到国内外计算机专家和学者的普遍关注。现在,许多高校都在积极开展计算思维的教学研究,重新审视与梳理计算机技术在各学科中的应用,培养学生的计算思维能力。

本教材第一版出版后受到了广泛的关注。许多教师、学生和读者在支持鼓励的同时,对本教材提出了许多宝贵的意见和建议。我们对这些意见和建议进行了认真的分析,提出了改进的思路,在出版社的大力支持下,经过努力完成了本教材的修订任务。在此修订版发行之际,我们对关爱、支持我们的各界人士表示衷心的感谢。

这次是对本教材进行的第一次修订,整体上基本保留了原书的结构,但对内容进行了更加适当的筛选,同时对例子和习题进行了大幅更新。各章节的具体变化如下:

第1章概述:将附录改为 Excel 数据分析工具。

第2章问卷设计与数据收集:附录中增加一个问卷实例(问卷实例二,大学入学新生信息技术与计算机基础情况调查问卷)。

第3章问卷数据的录入与清理:删去了“在 SPSS 中核对问卷数据”一节,将附录中的“Excel 数据分析工具”提前至第1章。

第4章单变量的一维频率分析:调整 4.3 节和 4.4 节的顺序,删去了“根据频率排名”一节,将其内容调整到 4.1.3 小节和 4.2.2 小节。

第5章双变量的交叉表分析:将原来在“利用 SPSS 实现两个单选题的交叉表分析”一节中的“在 Excel 中绘制百分比堆积柱形图”分离出来,增加了“在 Excel 中绘制两个单选题的交叉表统计图”一节。删去了附录中的“关于计算机课程教学情况调查问卷”。

第6章多选变量的一维频率分析和交叉表分析:将原来的“利用 SPSS 对多选题进行频率分析”一节拆分成三节,分别为:利用 SPSS 实现“二分法”编码多选题的一维频率分析、利用 SPSS 实现“分类法”编码多选题的一维频率分析和利用 SPSS 实现多选题的交叉表分析。将原来的“利用 Excel 对多选题进行一维频率分析”一节全部重新改写并拆分成两节,分别为:利用 Excel 实现“分类法”编码多选题的一维频率分析和利用 Excel 实现“二分法”编码多选题的一维频率分析。增加了一节:利用 Excel 实现“分类法”编码多选题的交叉表分析。删去了“撰写多选题的一维频率分析报告”一节,将其内容调整到相应的小节。

第7章描述统计分析:将原来在“利用 SPSS 实现多组均值比较”一节中的“利用 SPSS 实现定序变量的描述统计分析”分离出来,增加为两节:利用 SPSS 实现有序变

量的描述统计分析和利用 SPSS 实现有序变量的多组均值比较。删去了“利用 Excel 对定量变量进行描述统计分析”和“利用 Excel 求量表均值并排名”两节，增加一节：利用 Excel “描述统计”分析工具实现矩阵题的描述统计分析。删去了附录：简化版的“手机营销组合”调查问卷。

第 8 章简单统计推断：假设检验：删去了“总体比例的检验”一节。

第 9 章单因素方差分析和第 10 章线性相关分析与线性回归分析改动不大。

本教材的写作基础是安装于 Windows 7 操作系统上的中文版 SPSS 20 和中文版 Excel 2010。为了能顺利学习本教材介绍的例子，建议读者在中文版 SPSS 20 和中文版 Excel 2010 的环境下学习。

为了使广大读者更好地掌握本教材的内容，加深理解并增强处理实际问题的能力，我们将本教材所有例题和习题的数据文件放在中国人民大学出版社的网站（[www.crup.com.cn/jingji](http://www.crup.com.cn/jingji)）上，读者可以登录该网站免费下载；为支持教师的教学，编著者还把她们多年教学中积累的教学课件奉献给教师们，需要的教师请与作者（[yexiang@ruc.edu.cn](mailto:yexiang@ruc.edu.cn)）或本教材的编辑（[2001liudong@163.com](mailto:2001liudong@163.com)）联系。

为方便教师教学和学生自学，我们还将出版本教材的同步配套辅导书《统计数据分析基础教程（第二版）习题与实验指导》。

本教材的修订由叶向和李亚平两位老师完成，她们是中国人民大学全校共同课“SPSS 基础与应用”课程的主讲教师，也是中国人民大学从事计算机基础教学工作的骨干教师。本教材修订的大部分内容，在“SPSS 基础与应用”课程及其他课程中讲授过多次，受到普遍欢迎。

在本教材的修订过程中参考了大量的国内外有关文献书籍，它们对本教材的成文起了重要作用。在此对一切给予支持和帮助的家人、朋友、同事、有关人员以及参考文献的作者一并表示衷心的感谢。

同时，也要感谢中国人民大学出版社的编辑，他们对本教材写作的支持以及对书稿的认真编辑和颇有效率的工作，使得本教材能尽快与读者见面。

鉴于编著者的水平和经验有限，教材第二版中仍难免有不当或失误之处，恳请各位专家和广大读者给予指正并提出宝贵意见，同时欢迎同行进行交流。编著者联系邮箱是：[yexiang@ruc.edu.cn](mailto:yexiang@ruc.edu.cn)。

最后，再次感谢多年来阅读和使用本教材的老师、同学和读者，感谢他们对本教材修改提出的宝贵意见和建议。

叶向

于中国人民大学信息学院

2014 年 12 月



# 第一版前言

在经济全球化进程不断加快、世界经济联系日趋紧密、市场竞争越来越激烈的今天，一个企业要想赢得市场，求得生存和发展，必须最大限度地减少决策失误的概率。为此，决策者仅凭个人的经验、知识和感觉是很难做到这一点的。在决策过程中，必须充分利用集体的经验、知识、智慧和科学的分析方法，对收集到的数据做出准确、及时的分析并制定正确的决策。掌握数据分析方法和实施工具，是现代管理人才必备的基本技能。

近年来在西方发达国家，信息技术人才和统计应用人才一直排名在就业需求榜的前列，具备计算机知识和统计知识的复合型人才在未来将具有巨大的发展前景和明显的从业优势。

本教材坚持以案例为依托，利用国内外普遍流行的 SPSS 统计分析软件以及最普遍流行的 Excel 软件来解决案例中的问题，使读者能更好地利用数据分析方法和实施工具解决实际问题，使数据分析方法在决策中能发挥重要作用，也使学生对统计应用更加感兴趣。

本教材的主要内容包括：问卷设计及数据收集、问卷数据录入与清理、问卷数据基本统计分析（单变量的频率分析、双变量的交叉表分析、多选变量的频率分析、描述统计分析）、假设检验、单因素方差分析以及相关与回归分析。

笔者从事统计应用教学与研究十多年，收集整理了丰富的案例和实践经验。本教材重点突出以下特点：

1. 把 SPSS 和 Excel 放在一起介绍。本教材的写作基础是安装于 Windows XP 操作系统上的 SPSS 13.0 英文版和 Excel 2003 中文版。

2. 对统计分析方法的介绍，力求通俗易懂、简明扼要。

3. 应用实际案例，从实际问题出发，重点介绍软件是如何帮助解决实际问题的，并不强调对软件中每个细节的介绍。

4. 在介绍应用 SPSS 和 Excel 软件进行统计分析时，给出了详细的步骤。对输出结果也尽量以读者较容易接受的口语方式进行阐述，而不是用难懂的统计术语讲解。也就是说，过程和说明并重，告诉读者如何根据统计分析结果撰写调查报告。

5. 每章提供了实际案例及相关数据，供读者练习。

本教材的大部分内容，在中国人民大学计算机应用类课程《SPSS 基础与应用》及

其他课程中讲授过多次,受到普遍欢迎。其教学辅助资源(习题与实验指导、课件、考题等)也在同步建设中。

经过整整8个月紧张的写作,终于在祖国60周年华诞的喜庆日子里完成了全部书稿。在本书出版之际,要感谢的人很多。首先要感谢信息学院信息系的陈禹教授和方美琪教授,让笔者有机会于1998年3—11月到香港理工大学计算机系进行有关数据仓库与数据挖掘方面的合作研究。1999年2月,陈禹老师让笔者跟李丘副教授一起给全校本科生讲授《统计分析软件SPSS》,从李丘老师那里学到很多,特别是他能够利用统计分析软件解决实际问题(从调查问卷设计、在Excel中录入数据、在SPSS中进行统计分析、在Excel中绘制图表到在Word中撰写调查报告的整个社会调查过程)的讲课方法,给笔者留下了深刻印象。让笔者明白了,学生需要我们能够教给他们解决实际问题的能力。2002年2—6月,笔者去听了李燕琛副教授退休前最后一次给全校硕士研究生开设的《统计分析软件SPSS》和《数据分析软件——Excel高级功能》两门课程。在李老师的课堂上,笔者开始感受到了Excel强大的数据处理与分析功能。2003年8—12月,劳动人事学院的潘锦棠教授给了我一次组织大型的全国性的调查问卷数据录入和统计分析的实践机会。2003年9月,陈禹教授让笔者给信息学院的硕士研究生开设方法课《现代统计方法》,于是笔者去听了统计学院吴喜之教授给全校博士生开设的方法课《统计模型及应用》。笔者非常喜欢吴喜之老师的这门课,从头到尾认真真听了两遍。从吴老师的课中,笔者知道了统计课程还可以这么教:“统计已经渗入到人们的社会、生活、工作等各个领域;以应用为目标学习统计,通过学习获得解决和处理问题的能力;要还统计应用以其本来面目,使得统计变成人人都能够基本上理解和掌握的有用工具”。2009年3月,笔者正式从信息系调入信息技术基础教研室,杨小平教授和尤晓东副教授让我负责《SPSS基础与应用》这门课,包括教学内容、教材编写、教学系统、课件、考题等。要感谢这些给我机会的老师们、使我重新感受到做学生的幸福的教授们以及让我“教学相长”的学生们。

这里还要特别感谢策划本书的中国人民大学出版社的潘旭燕老师,她非常热心,工作认真负责,一直鼓励我将有自己特色的统计应用教学方法写出来,也告诉我很多把书写好的方法。在编写过程中参考了大量的国内外有关文献书籍,它们对本书的成文起了重要作用。在此对一切给予支持和帮助的家人、朋友、同事、同学、有关人员以及参考文献书籍的作者一并表示衷心感谢。

为了使广大读者更好地掌握本教材的内容,加深理解并增强处理问题的能力,我们将本书所有例题和习题的数据文件放在中国人民大学出版社的网站([www.crup.com.cn](http://www.crup.com.cn))的资源中心处,读者可以登录该网站免费下载;为支持教师的教学,本书的作者还将她多年教学中积累的教学课件奉献给老师们。需要的老师,请与本书作者或中国人民大学出版社编辑部联系,电子邮箱:[yexiang@ruc.edu.cn](mailto:yexiang@ruc.edu.cn)或[panxuyan@263.net](mailto:panxuyan@263.net)。

为方便教师教学和学生自学,我们还将出版本教材的配套辅导书《统计数据分析基础教程习题与实验指导》。

鉴于编著者的水平和经验有限,书中错误和不妥之处在所难免,恳请各位专家和

广大读者给予指正并提出宝贵意见，同时欢迎同行进行交流。编著者联系邮箱是：  
yexiang@ruc.edu.cn。

**叶向**  
于中国人民大学信息学院  
2009年10月

# 目 录

## CONTENTS

---

<b>第 1 章 概述</b> .....	1
1.1 什么是统计 .....	1
1.2 统计、计算机与统计软件 .....	2
1.3 为何要使用 Excel 来学习统计 .....	3
1.4 变量及其分类 .....	4
1.5 数据的收集 .....	6
1.6 思考题与上机实验题 .....	10
本章附录 Excel 数据分析工具 .....	10
<b>第 2 章 问卷设计与数据收集</b> .....	13
2.1 问卷的概念及其结构 .....	13
2.2 设计问卷的步骤 .....	16
2.3 几种典型的问卷题型 .....	18
2.4 “态度 8” 问卷模板库简介 .....	25
2.5 编辑问卷的技巧 .....	27
2.6 收集问卷数据 .....	29
2.7 思考题与上机实验题 .....	34
本章附录 I 问卷实例一 .....	35
本章附录 II 问卷实例二 .....	38
本章附录 III 调查研究方案实例 .....	41

<b>第 3 章 问卷数据的录入与清理</b> .....	45
3.1 问卷数据的录入 .....	45
3.2 在 Excel 中录入问卷数据 .....	47
3.3 核对和清理问卷数据 .....	55
3.4 在 Excel 中核对问卷数据 .....	58
3.5 建立调查问卷的 SPSS 数据文件 .....	67
3.6 思考题与上机实验题 .....	78
本章附录 在 Excel 2010 中生成随机数 .....	81
<b>第 4 章 单变量的一维频率分析</b> .....	89
4.1 利用 SPSS 实现单选题的一维频率分析 .....	89
4.2 利用 Excel 实现单选题的一维频率分析 .....	103
4.3 在 Excel 中绘制单选题的一维频率分布统计图 .....	114
4.4 如何用 Word 编辑一维频率分布表和统计图 .....	124
4.5 利用 SPSS 实现填空题的一维频率分析 .....	128
4.6 利用 Excel 实现填空题的一维频率分析 .....	134
4.7 撰写调查报告 .....	139
4.8 思考题与上机实验题 .....	145
本章附录 社会调查报告实例 (频率分析) .....	145
<b>第 5 章 双变量的交叉表分析</b> .....	149
5.1 利用 SPSS 实现两个单选题的交叉表分析 .....	149
5.2 在 Excel 中绘制两个单选题的交叉表统计图 .....	157
5.3 利用 Excel 数据透视表实现单选题的一维频率分析和交叉表分析 .....	166
5.4 交叉表行列变量间关系的分析 .....	186
5.5 思考题与上机实验题 .....	189
本章附录 社会调查报告实例 (交叉表分析) .....	190
<b>第 6 章 多选变量的一维频率分析和交叉表分析</b> .....	198
6.1 利用 SPSS 实现“二分法”编码多选题的一维频率分析 .....	198
6.2 利用 SPSS 实现“分类法”编码多选题的一维频率分析 .....	209
6.3 利用 SPSS 实现多选题的交叉表分析 .....	213
6.4 在 Excel 中绘制多选题的一维频率分布条形图和交叉表簇状条形图 .....	221
6.5 利用 Excel 实现“分类法”编码多选题的一维频率分析 .....	231
6.6 利用 Excel 实现“二分法”编码多选题的一维频率分析 .....	237
6.7 利用 Excel 实现“分类法”编码多选题的交叉表分析 .....	242
6.8 思考题与上机实验题 .....	248

<b>第 7 章 描述统计分析</b> .....	250
7.1 利用 SPSS 实现定量变量的描述统计分析 .....	250
7.2 利用 SPSS 实现定量变量的多组均值比较 .....	255
7.3 利用 SPSS 实现有序变量的描述统计分析 .....	262
7.4 利用 SPSS 实现有序变量的多组均值比较 .....	269
7.5 利用 Excel “描述统计” 分析工具实现矩阵题的描述统计分析 .....	279
7.6 思考题与上机实验题 .....	284
<b>第 8 章 简单统计推断：假设检验</b> .....	285
8.1 假设检验的基本原理 .....	285
8.2 利用 SPSS 实现单样本 t 检验 .....	289
8.3 利用 SPSS 实现独立样本 t 检验 .....	293
8.4 利用 SPSS 实现配对样本 t 检验 .....	296
8.5 利用 Excel 实现单样本 t 检验 .....	301
8.6 利用 Excel 实现独立样本 t 检验 .....	303
8.7 利用 Excel 实现配对样本 t 检验 .....	306
8.8 思考题与上机实验题 .....	308
<b>第 9 章 单因素方差分析</b> .....	311
9.1 单因素方差分析的基本原理 .....	311
9.2 利用 SPSS 实现单因素方差分析 .....	313
9.3 利用 Excel 实现单因素方差分析 .....	318
9.4 思考题与上机实验题 .....	320
<b>第 10 章 线性相关分析与线性回归分析</b> .....	321
10.1 问题的提出 .....	321
10.2 定量变量的线性相关分析 .....	322
10.3 利用 SPSS 实现线性相关分析 .....	323
10.4 定量变量的线性回归分析 .....	324
10.5 利用 SPSS 实现线性回归分析 .....	326
10.6 利用 Excel “图表” 实现一元线性回归分析 .....	329
10.7 利用 Excel “回归” 分析工具实现多元线性回归分析 .....	334
10.8 思考题与上机实验题 .....	337
<b>参考文献</b> .....	340

# 第 1 章

## 概 述

本章将介绍统计数据分析中经常用到的一些基本概念，包括什么是统计、统计与计算机的关系、统计软件、变量、数据的收集等内容。

### 1.1 什么是统计

你想过下面的问题吗?<sup>①</sup>

(1) 当你买了一台电视，被告知三年内可以免费保修时，你想过厂家凭什么这样说吗？说多了，厂家会损失；说少了，会失去竞争，也是损失。到底这个保修期是怎样决定的呢？

(2) 在同一年级中，同一门统计学的课程可能由一些不同的教师讲授。教师讲课方式当然不一样，考试题目也不一定相同。那么如何比较不同班级的统计学成绩呢？

(3) 大学排名是一个非常敏感的问题。不同的机构会得出不同的结果，各自都说自己是客观、公正和有道理的。到底如何理解这些不同的结果呢？

(4) 如何通过大众调查来得到性别、年龄、职业、收入等各种因素与公众对某件事物（比如商品或政策）的态度的关系呢？

(5) 如何才能够客观地得知某个电视节目的收视率，以确定广告的价格是否合理呢？

其实，这些都是统计应用的例子。这样的例子太多了。因为统计学可以应用于几乎所有的领域，包括社会学、新闻调查、精算、农业、动物学、人类学、考古学、审计学、人口统计学、牙医学、生态学、计量经济学、教育学、选举预测和策划、工程、

<sup>①</sup> 参见吴喜之编著：《统计学：从数据到结论》（第二版），1页，北京，中国统计出版社，2006。

流行病学、金融、水产渔业研究、遗传学、地理学、地质学、历史研究、人类遗传学、水文学、工业、法律、语言学、文学、劳动力计划、管理科学、市场营销学、医学诊断、气象学、军事科学、眼科学、制药学、物理学、政治学、心理学、心理物理学、质量控制、宗教研究、分类学、气象改善、博彩等。当然，大家用不着也不可能理解所有的统计应用，只要能够解决自己身边的统计问题就足够了。

上面的例子并没有明确说出什么是统计。其实很简单，上面的所有例子都要通过各种直接或间接的手段来收集数据 (Data)，都要利用一些方法来整理和分析数据，最后通过分析得到结论。一句话，统计学 (Statistics) 是用以收集数据、分析数据并进而由数据得出结论的一组概念、原则和方法。因而有学者也将统计学称为统计方法 (Statistical Method)。比如，要得到某电视节目的收视率，可能首先要在该节目播出时，利用电话对看电视的人进行采访，同时问他们在观看什么节目。在得到了被采访的看电视的总人数和其中观看该节目的人数之后，就有可能得到这部分观众中观看该节目的比例，即大致的收视率了。之后还要经过统计分析，评估这个收视率的可信度和代表性等。显然，这是一个收集数据，然后通过分析数据得到结论的简单例子。

## 1.2 统计、计算机与统计软件

现代生活越来越离不开计算机了。最早使用计算机的统计当然更离不开计算机了。事实上，最初的计算机仅仅是为科学计算而设计和制造的。大型计算机最早的一批用户就包含统计。现在，统计仍然是进行数字计算最多的用户。当然，计算机现在早已脱离了仅有数字计算功能的单一模式，成为了百姓生活的一部分。计算机的使用，也从过去必须学会计算机语言发展到只需要“傻瓜式”地点击鼠标；结果也从单纯的数字输出发展到包括漂亮的表格和图形在内的各种形式。<sup>①</sup>

统计软件的发展，也使得统计从统计学家的圈内游戏变成了大众的游戏。只要输入你的数据，点几下鼠标，做一些选项，马上就得到令人惊叹的漂亮结果了。人们可能会问，是否傻瓜式统计软件的使用可以代替统计课程？当然不是。数据的整理和识别，方法的选用，计算机输出结果的理解都不像使用傻瓜相机那样简单可靠。有些诸如法律和医学方面的软件都有不少警告，不时提醒你去咨询专家。但统计软件则不那么负责，只要数据格式无误、选项不矛盾而且不用零作为除数就一定给你结果，而且几乎没有任何警告。另外，统计软件输出的结果太多。即使是同样的方法，不同软件输出的内容也不一样。有时同样的内容名称也不一样。这就使得使用者大伤脑筋。即使是统计学家也不一定能解释所有的输出。因此，就应该特别留神，明白自己是在干什么，不要在得到一堆毫无意义的垃圾之后还沾沾自喜。

统计软件的种类很多。有些功能齐全，有些价格便宜，有些容易操作，有些需要更多的实践才能掌握。还有些是专门的软件，只处理某一类统计问题。面对太多的选

<sup>①</sup> 参见吴喜之编著：《统计学：从数据到结论》(第二版)，8~9页，北京，中国统计出版社，2006。



择往往给决策带来困难。这里介绍最常见的几种。

### 1. SPSS

这是一个很受欢迎的统计软件。它操作容易，输出漂亮，功能齐全，价格合理。它也有自己的程序语言，但基本上已经“傻瓜化”。对于非专业统计工作者，它是很好的选择。

### 2. Excel

严格说来，Excel 并不是统计软件，但作为数据表格软件，必然有一定的统计计算功能。而且凡是安装了 Microsoft Office 的计算机，基本上都装有 Excel。但要注意，如果在安装 Office 时没有安装（加载）“数据分析”的功能，那么必须安装（加载）才行。<sup>①</sup>当然，画图功能是已经具备了。对于简单分析，Excel 还算方便，但随着问题的深入，Excel 就不那么“傻瓜”了，需要使用宏命令来编程，这时就没有相应的简单选项了。多数较为专业的统计推断问题还需要其他专门的统计软件来处理。

### 3. SAS

这是一款功能非常齐全的软件。尽管价格相当不菲，但许多公司，特别是美国制药公司都在使用，这多半因为其功能众多和某些美国政府机构中一些人的偏爱。尽管现在已经尽量“傻瓜化”，但仍然需要一定的训练才可以使用。也可以用它编程计算，但对于基本统计课程则不那么方便。

当然，还有很多其他的软件，如 S-Plus、R 软件、MATLAB 等，这里不一一罗列。其实，聪明的读者只要学会使用一种“傻瓜式”软件，使用其他的软件也不会困难；最多看看帮助和说明即可。如果只有英文帮助，还可以顺便提高英文阅读能力。学习软件最好的方式是需要时在使用中学。

## 1.3 为何要使用 Excel 来学习统计

前面介绍的 SPSS、SAS 等统计分析软件，在市面上的普及率很低，在个人计算机上，占有率甚至连千分之一都不到。<sup>②</sup>其原因如下：

### 1. 价格昂贵

这几种软件价格高达上万到几十万。通常，教育单位、国家机关或大型研究单位才有能力购买；一般学生、个人或中小企业都因负担不起而不可能购买。

### 2. 学习困难

这几个软件，若仅是要进行操作而取得分析结果，在有人指导下或有相关优秀书籍来参考的情况下并不困难。但因学习的人不多，受过完整训练的师资本来就不多，市面上可用的书籍也不是很多。所以，对大部分人来说，学习起来还是很困难的。况且很多单位所使用的还是英文版，更加大了学习的难度。

<sup>①</sup> 参见本章附录。

<sup>②</sup> 参见杨世莹编著：《Excel 数据统计与分析范例应用》，2 页，北京，中国青年出版社，2008。