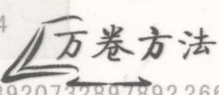


0.02\32156465648765497008.70082635.6890.7892073289032156465648765497008.70082635.6890.20
6465648765497008.70082635.6890.7892073289032156465648765497008.70082635.6890.20
988944564...0.02\32156465648765497008.70082635.6890.789207328990.7892073289789220732890.
5.6890564 32890...59744132667988944564...0.02\32156465648765497008.7008263
44564... 54765497008.70082635.6890.7892073289789220732890...597441326679889
328978 78920732897892266732667984132667988944564...0.02\32156465648765497008.70082635.68
07328...597 03213266798.0321326679890.7892073289789220732890...5974413266798010508971003541.
/319645912.05932147850963214785120.15/225465189961132.0.85474165849897711223...15445574985
5497008.70082635.6890.7892073289032156465648765497008.70082635.6890.20732890...59744132667
2156465648765497008.70082635.6890.789207328990.7892073289789220732890...
220732890...59744132667988944564...0.02\32156465648765497008.70082635.
465648765497008.70082635.6890.7892073289789220732890...5974413266798
59744132667984132667988944564...0.02\32156465648765497008.70082635.689
597441326679890.7892073289789220732890...597441326679801050897100354164655434844656440908
20963214785120.15/225465189961132.0.85474165849897711223...15445574985214125696355477.
0.02\32156465648765497008.70082635.6890.7892073289032156465648765497008.70082635.6890.20
988944564...0.02\32156465648765497008.70082635.6890.789207328990.7892073289789220732890.
5.6890.78920732897892266732667984132667988944564...0.02\32156465648765497008.7008263
44564...0.02\32156465648765497008.70082635.6890.7892073289789220732890...597441326679889
3289789220732890...59744132667984132667988944564...0.02\32156465648765497008.70082635.68
073289789220732890...5974413266798010508971003541.
/31964598436412.0.0320963214785120.15/225465189961132.0.85474165849897711223...15445574985
/31964598436412.0.0320963214785120.15/225465189961132.0.85474165849897711223...15445574985
5497008.70082635.6890.7892073289032156465648765497008.70082635.6890.20732890...59744132667
2156465648765497008.70082635.6890.789207328990.7892073289789220732890...
220732890...59744132667988944564...0.02\32156465648765497008.70082635.689
465648765497008.70082635.6890.7892073289789220732890...59744132667988944564...0.02\32156
9744132667984132667988944564...0.02\32156465648765497008.70082635.6890.78920720737892207
597441326679890.7892073289789220732890...59744132667980105086890.7892076890.7892076890908
20963214785120.15/225465189961132.0.85474165849897711223...15165849897711223696355477.1223
0.02\32156465648765497008.70082635.6890.7892073289032156465648765497008.70082635.6890.20
98894 32890...59744132667988944564...0.02\32156465648765497008.70082635.689
5.68 $H_0: \mu_9 = \mu$ 1232890...59744132667988944564...0.02\32156465648765497008.7008263
4456[3215892207 64 56465648765497008.70082635.6890.7892073289789220732890...597441326679889
3441326679841326...59744132667984132667988944564...0.02\32156465648765497008.70082635.68
073289789 32890...597441326679890.7892073289789220732890...5974413266798010508971003541.
1964598436412.0.0320963214785120.15/225465189961132.0.85474165849897711223...1544557498
11964598436412.0.0320963214785120.15/225465189961132.0.85474165849897711223...15445574985
97008.70082635.6890.78920732890321564656 654 8.70082635.6890.20732890...59744132667
5646564876549700 0920920 92073289789220732890...597441326679890.78
0732890... 564 656 5648765497008.70082635.6890.789207328990.
564876549 929220732890...59744132667988944564...0.02\32156
441326679 156465648765497008.70082635.6890.78920732897892207
974413266 732 ...597441326679801050897100354164655434844656440908
963214785120.15/225465189961132.0.85474165849897711223...15445574985214125696355477.
02\3215643289789220732890...2635.6890.7892073289032156465648765497008.70082635.6890.20
3944564...032156432156465648765497008.70082635.6890.789207328990.7892073289
5.6890.7892073289789220732890...59744132667988944564...0.02\32156465648765497008.7646564
44564...0.02\32156465648765497008.70082635.6890.7892073289789220732890...597441.78922073
3289789220732890...59744132667984132667988944564...0.02\32156465648765497008.70082635.68
073289789220732890...597441326679890.7892073289789220732890...5974413266798010508971003541
/319645984364120.0320963214785120.15/225465189961132.0.85474165849897711223...15445574985
073289789220732890...597441326679890.7892073289789220732890...5974413266798010508971003541

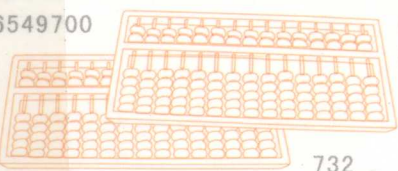


STATISTICS FOR THEM WHO (THINK THEY) HATE STATISTICS

爱上统计

(美) 尼尔·J·萨尔金德 著
史玲玲 译

$$H_0: \mu_9 = \mu$$



清华大学出版社
<http://www.cqun.com.cn>

万卷方法

C8/176

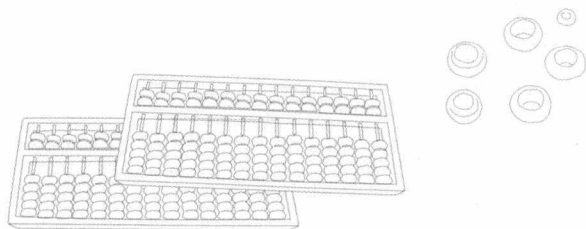
2008

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

STATISTICS FOR THEM WHO (THINK THEY) HATE STATISTICS

爱上统计学

(美)尼尔·J·萨尔金德 著
史玲玲 译



重庆大学出版社

Authorized translation from the English language edition, entitled STATISTICS FOR PEOPLE WHO (THINK THEY) HATE STATISTICS, 2nd edition by Neil J. Salkind, published by Sage Publications, Inc., Copyright © 2005 by Sage Publications, Inc.

All rights reserved, No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. CHINESE SIMPLIFIED language edition published by CHONGQING UNIVERSITY PRESS, Copyright © 2006 by Chongqing University Press.

《爱上统计学》原书英文版由 Sage 出版公司出版。原书版权属 Sage 出版公司。

本书简体中文版专有出版权由 Sage 出版公司授予重庆大学出版社, 未经出版者书面许可, 不得以任何形式复制。

版贸渝核字(2006)第13号

图书在版编目(CIP)数据

爱上统计学/(美)萨尔金德(Salkind, N. J.)著;
史玲玲译. —重庆:重庆大学出版社, 2008. 1

(万卷方法)

书名原文: Statistics for People Who(Think They)

Hate Statistics

ISBN 978-7-5624-4196-0

I. 爱… II. ①萨…②史… III. 统计学—研究 IV. C8

中国版本图书馆 CIP 数据核字(2007)第 096008 号

爱上统计学

(第2版)

尼尔·J·萨尔金德 著

史玲玲 译

责任编辑: 雷少波 林佳木 版式设计: 雷少波

责任校对: 邹 忌 责任印制: 张 策

*

重庆大学出版社出版发行

出版人: 张鸽盛

社址: 重庆市沙坪坝正街174号重庆大学(A区)内

邮编: 400030

电话: (023) 65102378 65105781

传真: (023) 65103686 65105565

网址: <http://www.cqup.com.cn>

邮箱: fxk@cqup.com.cn (市场营销部)

全国新华书店经销

重庆华林天美印务有限公司印刷

*

开本: 787×1092 1/16 印张: 17.5 字数: 335千 插页: 16开2页

2008年1月第1版 2008年1月第1次印刷

印数: 1—4 000

ISBN 978-7-5624-4196-0 定价: 45.00元

本书如有印刷、装订等质量问题, 本社负责调换

版权所有, 请勿擅自翻印和用本书

制作各类出版物及配套用书, 违者必究

作者简介

尼尔·J·萨尔金德(Neil J. Salkind) 他在堪萨斯大学心理学和教育学系从教30年,同时兼任人类发展和家庭生活系的教授,30年来,开设的课程有发展理论、生命全程发展理论、统计学和研究方法。他获得马里兰大学人类发展的博士学位,发表了80多篇专业文章,也是一些大学教材的作者,如《儿童发展、探索性研究和人类发展理论导论》(Sage,2004)。他也是1989—2002年的《儿童发展摘要和书目》的编辑。他在儿童发展研究领域很活跃,在贸易领域也有很多著述。他住在堪萨斯州劳伦斯市一所需要经常维护的老房子里,喜欢烹饪、游泳、阅读、藏书及收藏古典沃尔沃P1800。

给学生的话：前作者 我为什么写这本书

我很幸运，三十年来能够在堪萨斯大学给成千上万的学生讲授基础统计学。他们中的大多数（至少在这门课开始时）都相当的焦虑，焦虑的原因通常是他们从其他学生那里听来的经验之谈。一般来说，他们听到的一小部分是正确的——学习统计学要投入很多的时间和精力（而且偶尔还会遇到“怪物型”老师）。但是他们听到的大部分——统计学特别难学，令人困扰——是不正确的，这恰恰是他们产生焦虑的原因。而实际上许多被吓倒的学生都通过了他们以为不能通过的课程。只要集中精力，按部就班，将基本原理应用于真实的生活来理解，他们都能通过这门课程，甚至能够在这个过程中得到乐趣。这就是我在写《爱上统计学》第1版的时候想尽力做到的，在这一版的修订中我更加努力做到这一点。

在经过不断地摸索以及少量成功大量失败的尝试之后，我已经学会了以某种方式教授统计学，我和我的许多学生认为这种方式不会让人感到害怕，同时能够传递大量的信息。

通过这本书可以了解基础统计学的范围并学习所有应该掌握的信息，也可以了解整理和分析数据的基本思路和最常用的技术。本书理论部分有一些，但是很少，数学证明或特定数学程式的合理性讨论也很少。

为什么《爱上统计学》这本书不增加更多理论内容？很简单，初学者不需要。这并不是我认为理论不重要，而是在学习的这个阶段，我想提供的是我认为通过一定程度的努力可以理解和掌握的资料，同时又不会让你感到害怕而放弃将来选修更多的课程。我和其他老师都希望你能成功。

因此，如果你想详细了解方差分析中F值的含义，可以从Sage出版社查找其他的好书（我愿意向你推荐书目）。但是如果你想了解统计学为什么以及如何为你所用，这本书很合适。这本书能帮助你理解在专业文章中看到的资料，解释许多统计分析结果的意义，并且能教你运用基本的统计过程。

祝大家好运，希望你们能让我知道如何修订这本书才能更好地满足初学统计学的学生的需求。

致 谢

一本好书是好的编辑和作者共同努力的结果。如果这本书真的实现了最初的设想，也是由于C·德博拉·劳顿(C. Deborah laughton)的指导、建议、努力工作和大力支持，并且使得这本书最终能够成书。在德博拉·劳顿离开Sage之后，爱丽森·马蒂特(Alison Mudditt)与利萨·奎瓦斯-肖(Lisa Cuevas-Shaw)承担了编辑的责任，而且他

们总是在需要的时候给予帮助。帕姆·格林(Pam Ggreen)阅读了本书的初稿,而且帮我找出疏漏之处,尽管我确信那不是全部,但是他的工作对本书质量的提高是极大的帮助。同时感谢文字加工编辑里安·列赫(Liann Lech)、制作编辑戴安娜·阿克塞尔森(Diana Axelsen)、内页设计和排版员提姆·吉森(Tim Giesen)、艺术制作桑德拉·额·索瓦约特(Sandra Ng Sauvajot)、卡通设计贾森·洛夫(Jason Love)以及封面设计米歇尔·李·肯尼(Michelle Lee Kenny)。

最后,与通常一样感谢一位具有非凡魅力的女性——莱尼(Leni)。

现在,关于第2版……

以上的内容说明了我写这本书的最初想法,但是关于新的版本介绍很少。

任何一本书都需要不断修订,《爱上统计学》也不例外。过去的三年许多人曾告诉我这本书是多么的有帮助,但是其他人也告诉我他们多么想让这本书做出修订以及为什么。在修订的过程中我尽力满足这两方面人的需求。书中部分内容仍保留,部分做出了修订。

总是有新的事物值得关注,同时也需要用不同的方式重新理解老的主题和观念。以下的清单是《爱上统计学》中新的内容。

- 第9章新增了效果值的简要介绍,这部分在第一版编辑过程中被删减了。新增的目的是让学生对差异的重要性有更深入的认识。
- 你会发现新增了第12章——析因方差分析。虽然第一版的一维分析对一些人已经足够,但是新增的这一章可进一步满足其他人更多的需要。
- 研究设计和统计分析中总是要考虑测量结果的工具和有效性检验。考虑到这一点,第16章增加了信度和效度的重要形式和相关问题的简要综述。
- 每一章后都增加了更多的练习题,许多习题可以用到本书附录C的数据、教师资源光盘的数据或 http://www.soe.ku.edu/faculty/salkind/stats_fpwhs/ 的在线数据。在线数据可以依据需要下载。
- 现在的数据库一般有两种形式——SPSS(流行的统计分析软件)和Excel(许多人使用电子表格分析数据)。
- 第1版中出现的所有错误都是我的责任,我向被这些错误困扰的老师和学生道歉。我非常感谢所有那些指正错误并使得第2版更好的信件、电话和电子邮件。在这一版,我们每个人都尽力修改错误,并希望工作做得更好。期望能收到大家的建议、批评和意见。祝大家好运。

尼尔·J·萨尔金德

第 I 部分 耶！我喜欢统计学

1 统计学还是虐待学？由你决定	4
为什么学习统计学	4
统计学简史	4
统计学：是什么（或不是什么）	6
我在统计学课堂上做什么	7
使用这本书的十种方式（同时也在学统计学！）	8
关于那些符号	10
难度指数	11

第 II 部分 西格玛·弗洛伊德和描述统计

2 必须完成的功课——计算和理解平均数	14
计算均值	14
需要记忆的内容	15
计算中位数	17
需要记忆的内容	19
计算众数	19
何时用什么	21
应用计算机并计算描述统计值	21
3 性别差异——理解变异性	26
为什么理解变异性很重要	26
计算极差	27
计算标准差	27
需要记忆的内容	31
计算方差	31
使用计算机计算变异性量数	32
4 一幅图真的相当于千言万语	35
为什么要用图表说明数据	35
好图表的十个方面（少贪新，多练习）	35

首先是建立频数分布	36
图形密度:建立直方图	37
扁平 and 细长的频数分布	43
其他的图表数据的绝妙方法	46
使用计算机图示数据	48
5 冰淇淋和犯罪——计算相关系数	56
相关系数到底是什么	56
需要记忆的内容	57
计算简单相关系数	58
理解相关系数的含义	64
决定性的努力:相关系数平方	64
其他重要的相关	66
使用计算机计算相关系数	67

第Ⅲ部分 抓住那些有趣又有利的机会

6 你和假设:检验你的问题	74
也许你想成为一个科学家	74
零假设	75
研究假设	77
好假设的标准是什么	80
7 你的曲线是正态的吗——概率和概率的重要性	84
为什么学习概率	84
正态曲线(或钟型曲线)	84
我们最中意的标准值: z 值	89
使用计算机计算 z 值	95

第Ⅳ部分 显著性差异——使用推论统计

8 显著性的显著——对你我来说意味着什么	100
显著性的概念	100
显著性与意义	105
推论统计介绍	106
显著性检验介绍	107
9 两个群体的 t 检验——不同群体的均值检验	112
独立样本 t 检验介绍	112
计算检验统计量	113
特殊效果:差异是真实的吗	117
使用计算机进行 t 检验	120
10 两个群体的 t 检验——两个相关群体的均值检验	125

非独立样本 t 检验介绍	125
计算检验统计量	126
使用计算机进行 t 检验	130
11 两个群体是否太多? ——尝试进行方差分析	134
方差分析介绍	134
计算检验统计量	137
使用计算机计算 F 比率	143
12 两个因素——析因方差分析	149
析因方差分析介绍	149
主要方面:析因方差分析的主效应	152
更有趣的方面:交互效应	153
需要记忆的内容	154
计算检验统计量	154
13 近亲还是好朋友——使用相关系数检验关系	160
相关系数检验的介绍	160
计算检验统计量	162
使用计算机计算相关系数	165
14 预测谁将赢得超级杯——使用线性回归	169
什么是估计	169
估计的逻辑	170
绘制拟合数据的最优直线	173
使用计算机计算回归线	176
估计变量越多就越好? 也许是	178
15 非正态分布时做什么——卡方和其他非参数检验	182
非参数统计的介绍	182
单样本卡方检验介绍	183
计算卡方检验统计量	183
使用计算机进行卡方检验	186
你应该了解的其他非参数检验	187
16 这就是真相——理解信度和效度	191
信度和效度介绍	191
关于测量尺度	192
信度——再做一次直到得到正确的值	194
使用计算机计算克隆巴赫系数	199
效度——哦! 真相是什么	202
信度和效度:很亲的堂表兄弟关系	205
17 你应该了解的其他重要的统计过程	207
多元方差分析	207

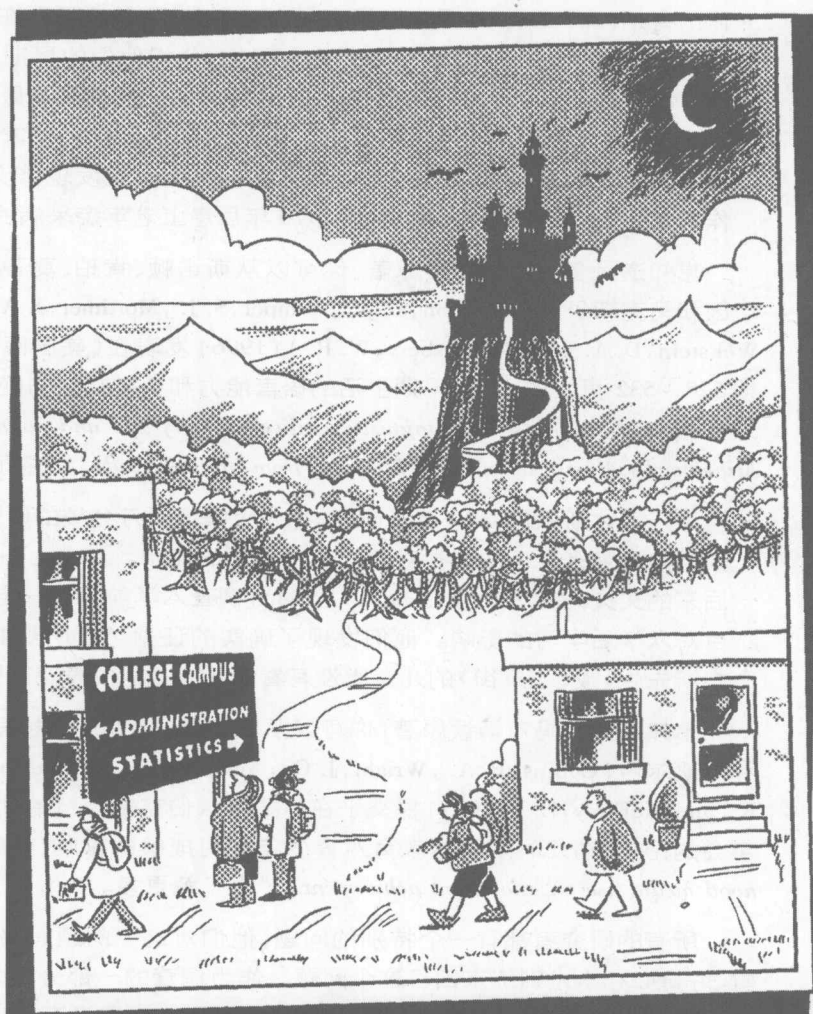
重复测量的方差分析	208
协方差分析	208
多元回归	209
因子分析	209
路径分析	210
结构方程模型	210
18 统计软件简介	212
选择合适的统计软件	212
具体介绍	213

第V部分 你得了解和记忆的内容

19 10个(或更多)最好的统计网址	220
成堆的资源	220
大量的计算器	221
谁是谁以及发生了什么	221
都在这里	221
超级统计(HyperStat)	222
数据,你想要数据	222
越来越多的资源	223
容易,但是有趣	223
在斯德哥尔摩学习统计学如何	223
越来越多的资源——再一次强调	224
在线统计学教学资料	224
越来越多的资料	224
20 收集数据的10个原则	225
附录A 30分钟SPSS教学	227
附录B 数据表	245
附录C 数据集	256

耶！ 我喜欢统计学

Yippee! I'm in Statistics



这有什么好欢呼的,你会这样说吗?现在给我几分钟的时间向你展示一些非常成功的科学家如何使用被广泛使用的叫做统计学的工具。

- 米歇尔·兰普尔(Michelle Lamp)是艾莫里大学的儿科专家和人类学家。她和朋友喝咖啡的时候,朋友谈到她的孩子长得多么多么的快。实际上,这个初为人母的朋友几乎是说她的儿子“像野草一样疯长”。兰普尔博士十分的好奇(像所有的科学家对事物感到好奇一样),她想她应该实际测量这个孩子以及其他孩子在婴儿期的生长速度。她着手测量一群孩子每天的生长,让她非常吃惊的是她发现一些婴儿差不多一个晚上就长一英寸!确实是爆发式生长。

想知道更多吗?请阅读原著:你可以从兰普尔、费尔德伊斯和约翰森(Lamp, M., Veldhuis, J. D., and Johnson, M. L.)(1992)发表在《科学》258期801~803页上的文章“突变和停滞:人类生长模式(*Saltation and stasis: A Model of human growth.*)”中了解更多。

- 休·肯珀(Sue Kemper)是堪萨斯大学心理学教授,曾经研究许多非常有趣的项目。她和其他的研究者正在研究修女群体,分析他们的早期经验、活动、人格特征和其他信息与他们中年以后的健康状况之间的关系。最特别的是这个由不同科学家组成的小组(包括心理学家、语言学家和神经学家,等等)想知道所有信息多大程度上可以预测老年痴呆病的发生。她发现修女在20多岁时写作的复杂性和他们在50、60年或者70年后患上老年痴呆病的可能性有关。

想知道更多吗?请阅读原著:你可以从斯诺顿、肯珀、莫蒂默、格雷纳、韦斯坦坦和马克斯贝里(Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. A., and Markesbery, W. R.)(1996)发表在《美国医学协会杂志》275期528~532页上的文章“早期生活的语言能力和认知功能与晚期生活的老年痴呆症:修女研究的发现(*Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the nun study.*)”中了解更多。

- 阿莱莎·休斯顿(Aletha Huston)是德州大学奥斯汀分校的研究者和教师,投入大量的研究来发现看电视对幼儿心理发展的作用。其中之一的工作是她和她后来的丈夫约翰·赖特(John. C. Wright)调查入学前观看一定量的电视教育节目对入学后学习的影响。他们发现了确实的证据可证明观看教育节目如《罗杰斯先生》或《芝麻街》的儿童比没有看的儿童在学校表现得更好。

想知道更多吗?请读原著:你可以从柯林斯、赖特、安德森、休斯顿、施密特和麦凯尔伊(Collins, P. A., Wright, J. C., Anderson, D. R., Huston, A. C., Schmitt, K. and McElroy, E.)(1997)提交于在美国阿尔伯克基举行的儿童发展研究协会年会的论文“幼儿时期媒体教育对青少年学习成绩的影响(*Effects of early childhood media use on adolescent achievement.*)”中了解更多。

所有的研究者都有一个特别的问题,他们对这个问题的兴趣并用他们的直觉、好奇心和学术训练来回答这个问题。作为调查的一部分,他们使用我们称作统计学的工具分析所收集到的所有资料的意义。如果没有这些工具,所有的资料就是不相关的资料的汇集。那么这些资料就不能在兰普尔的研究中用于得出

儿童生长的结论，不能在肯珀的研究中用于对老年痴呆病有更好的理解，也不能在休斯顿和赖特的研究中用于更好地理解看电视对幼儿学习和社会发展的影响。

统计学——整理和分析资料并使得资料更容易理解的科学——使得研究任务可行。

通过这样的研究所得出的任何结论都是有用的，原因是我们使用统计学使得这些结果有意义。这也正是本书的目标——让你理解这些基本的工具以及这些工具的用途，当然也包括如何使用这些工具。

在《爱上统计学》的第 I 部分介绍统计学学习的内容，以及为什么值得花费精力来掌握这个领域关键的基本术语和思路。这都是学习本书其他部分的准备。

统计学还是虐待学？由你决定

本章你会学到什么 ☺☺☺☺☺

- 统计学的学习内容
- 为什么学习统计学
- 如何通过这门课程的考试

为什么学习统计学

你以前肯定听到过“统计学很难学”，“统计中相关的数学部分很难应付”，“我不知道怎么用电脑”，“学统计学有什么用”，“接下来做什么”，还有就是统计学导论课上学生的著名呐喊，“我就是不懂！”

好啦，放松些。学习统计学导论的学生总会发现他们自己时不时有以上的想法，如果他们不和其他学生或者他们的爱人、同事或者朋友交流的话，他们至少有上面的一个想法。

而且，不是开玩笑，一些统计学课程很容易被描述为虐待学。这是由于那些书无一例外地让人厌烦，而且作者没有想象力。

本书将改变这种状况。事实是，你或者你的老师选择了《爱上统计学》，这表明你准备选择正确的方法——不会让人害怕、信息充分并且很实用（甚至有趣）的方法，而且尽可能地教你应用这一有价值的工具——统计学的确是有价值的工具，需要知道些什么。

如果选用这本书作为教材，这也意味着你的老师明显地站在你的立场上——他或她知道统计学是让人害怕的，但他们已经采取措施来确保统计学不会让你害怕。事实上，我敢打赌你在几个星期之后享受这门课程的机会很大（这同样也难以让人相信）。

统计学简史

在阅读更多内容之前，有必要从历史的视角了解统计学。你知道，毕

竟几乎所有的社会科学、行为科学、生物学的大学生和教育学、护理学、心理学、社会福利和社会服务以及人类学的研究生都需要选修这门课。了解统计学应用领域从何处开始是否有必要? 回答是当然有必要。

只要回顾过去,人们就会发现计数是个非常好的创意(就如“你需要多少这个来交换那些中的一个”),同样收集信息也成为有用的技能。如果用到计数,人们就会知道太阳在一季升起多少次,度过整个冬天需要多少食物,以及一个人拥有多少资源。

这只是开始。一旦数字成为语言的组成部分,似乎接下来的步骤就是将数字和结果结合在一起。在 17 世纪早期就开始收集与人口相关的数据。以此为基础,科学家(大多数是数学家,此外是物理学家和生物学家)需要发展特定的工具来回答特定的问题。例如,弗兰茨·加尔东(Francis Galton)(顺便说一下,他是达尔文的表兄弟,生卒年代是 1822—1911 年),他对人类智力的性质非常感兴趣。为回答家庭成员智力一致性的基本问题,他使用了特定的统计工具——相关系数(首先是数学家发展的),接着他把相关系数广泛地应用于行为科学和社会科学。你会在第五章对这个工具有全面的了解。

实际上,你将学到的大多数基本统计过程最初应用于农业、天文学甚至政治学领域。在人类行为领域的应用则比较晚。

在过去的 100 年中,在发明新方法应用旧观念方面取得了极大的进步。最简单的用于比较两个群体的均值差异的检验方法在 20 世纪初取得首次进展。在此基础上建立的技术十年之后才提出,而且得到了极大的完善。随着个人电脑和类似 SPSS 的软件(见附录 A)的应用,任何人想研究这些有意义问题都可以使用复杂的统计技术。

影响力深远的个人电脑的应用有利有弊。有利的方面是大多数统计分析不再需要使用巨大、昂贵的中央处理机。而价值不超过 1 000 美元的个人电脑就能满足 95% 的人 95% 的需求。另一方面,很少有大学生(如已经通过这门课的你的同学)会使用他们已有的旧数据,并且认为应用复杂的 SPSS 软件分析这些数据可以得到可靠、可信和有意义的结果——这是不正确的。你的老师可能会说“垃圾进,垃圾出”——如果你开始就没有使用可靠、可信的数据,那么分析这样的数据所得到的结果就既不可信也不可靠。

现在,不同领域——从司法、地球物理学到心理学——的统计分析人员发现他们基本上使用相同的技术来回答不同的问题。当然在资料收集方面有重要的差别,但是大体上来说,随着数据(大量的数据)收集所进行的分析(多个的分析),即使说有所谓的差别也是非常类似的。那么伦理问题呢? 这个课上会给你提供一定的工具来理解统计学如何应用于几乎任何一个学科中,帮你轻松搞定那些 3~4 个学分的课程。

如果你想更多地了解统计学历史,而且想按历史发展来了解,可以从

一个很好的地方开始,就是圣安赛姆学院的网络,网址是 <http://www.anselm.edu/hompage/jpitocch/biostatshist.html>; 或者 <http://www.stat.ucla.edu/history/> (加利福尼亚大学洛杉矶分校)。

统计学:是什么(或不是什么)

《爱上统计学》是一本关于基础统计学,以及在不同的情况下如何应用统计学分析和理解资料的书。

就一般意义而言,统计学是描述一系列可用于描述、整理和解释资料或数据的统计工具和技术。这些数据可能是特定数学课程的学生的考试得分、解决问题的速度、吃某种药感到不适的病人人数、世界系列大赛每一局出错的次数或者圣达菲高级餐厅晚餐的平均价格。

在所有的案例或者我们可以想到的更多的案例中,都需要收集、整理、汇总和解释数据。在本书的描述性统计部分你可以学到收集、整理和汇总数据。而在了解了推论性统计的用途之后就可以学会解释数据。

什么是描述统计

描述统计(descriptive statistics)常用于整理、描述所收集数据的特征。所收集的资料有时也称作数据集(data set)或者就叫数据(data)。

例如,下面所列出的就是22名大学生的姓名、主修专业和年龄。如果需要描述大学最流行的专业是什么,你可以使用描述统计来概括他们的选择(也称作众数)。在这个例子中最大众化的专业是心理学。如果你想知道学生的平均年龄,可以很容易的计算另一个描述统计值(也叫做均值)来确定这个变量。这两个简单的描述统计值常用于描述数据。就如我们对下面22个案例的分析一样,描述统计使我们能很好地描述大的数据集的特征。

姓名	专业	年龄	姓名	专业	年龄
Richard	教育学	19	Elizabeth	英语	21
Sara	心理学	18	Bill	心理学	22
Andrea	教育学	19	Hadley	心理学	23
Steven	心理学	21	Buffy	教育学	21
Jordan	教育学	20	Chip	教育学	19
Pam	教育学	24	Homer	心理学	18
Michael	心理学	21	Margaret	英语	22
Liz	心理学	19	Courtney	心理学	24
Nicole	化学	19	Leonard	心理学	21
Mike	护理学	20	Jeffrey	化学	18
Kent	历史	18	Emily	西班牙语	19

你瞧,这是多么简单! 要确定最常被选择的专业,就是要确定哪一个

专业出现次数最多。要确定平均年龄,就是将所有的年龄值加起来然后除以 22。你是对的——出现次数最多的专业是心理学(9 次),平均年龄是 20.3。瞧,没人插手——你就是统计分析师。

什么是推论统计

推论统计(inferential statistics)通常是(但并非总是)数据收集和汇总后的下一步。推论统计常利用较小群体的数据(如我们的 22 个学生构成的群体)来推论可能的较大群体(如艺术和科学院的所有学生)的特征。

这个较小的群体通常叫做样本(sample),是总体(population)的一部分或一个子集。例如,新泽西州纽瓦克的所有五年级学生构成一个总体(所有具备这些特征——正好是五年级而且住在纽瓦克——的人都包括在内),从中选取 150 人就构成一个样本。

咱们来看另外一个例子。作为新雇用的研究人员,你的市场代理要你确定多个名称中的哪一个最适合作薯片的新品牌,是薯片王,乐趣薯片,还是嚼嚼乐?作为专业统计分析人员(我知道我们现在超前了一点,但是保持信念),你需要选取一个薯片食用者小群体,这个群体可以代表所有的喜欢吃薯片的人,接着询问这个群体让他们告诉你最喜欢这三个名称中的哪一个。如果你做得很好,就可以很容易地将这个发现推论到更大的薯片食用者群体中。

或者,如果说你对某种疾病的最佳治疗方案感兴趣。你可以让一组试用一种新药,另一组食用安慰剂(或者大家知道的没有任何效果的其他丸剂),第三组什么药也不吃,然后来看结果是什么。好吧,你会发现大多数病人在没有接受任何治疗情况下会好转,而且机体自身会恢复!药物没有任何效果。那么依据你的实验结果,你可以将这些信息推论到更大的遭受这种疾病痛苦的病人群体。

换句话说……

统计学是帮助我们理解周围世界的工具。这是通过整理我们收集到的数据实现的,而且接着还可以让我们做出特定的推断,也就是怎样将那些数据的特征应用到新的情况。描述统计和推论统计可以一起发挥作用,使用哪一种、何时使用取决于你想要回答的问题。

我在统计学课堂上做什么

你选择使用这本书可能有多种原因。也许是选修了统计学导论课程,或者是在申请综合考试,甚至是在暑假预习这本书(恐怖!),并且想申请更高级的课程。

总之你都是学习统计学的学生,不论你是否参加课程结束后的考试,