

On Teaching Multivariate Statistical Analysis in Maritime Universities

Ke Zhang

College of Arts and Sciences, Shanghai Maritime University, Shanghai
Email: kezhang@shmtu.edu.cn

Received: Jul. 25th, 2019; accepted: Aug. 8th, 2019; published: Aug. 15th, 2019

Abstract

By analyzing the status quo and existing problems in teaching multivariate statistical analysis of maritime universities, we explore various teaching strategies to stimulate students' interest in learning this course and improve their ability to solve real-world problems via statistical methods.

Keywords

Multivariate Statistical Analysis, Teaching Mode, Vivid Teaching

航海类院校多元统计分析课程的教学思考

张 科

上海海事大学文理学院, 上海
Email: kezhang@shmtu.edu.cn

收稿日期: 2019年7月25日; 录用日期: 2019年8月8日; 发布日期: 2019年8月15日

摘 要

分析航海类院校多元统计分析课程的教学现状及存在问题, 探索多元统计分析课程的多样化教学模式, 提高学生的学习兴趣及运用统计方法解决现实问题的能力。

关键词

多元统计分析, 教学模式, 形象教学

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

多元统计分析是研究多元随机变量之间的相互关系以及变量内在的变化规律的理论和方法,具有社会科学和自然科学的双重特征。在统计学发展史上,多元统计分析的研究始于英国统计学家 John Wishart 在 1928 年发表的论文《多元正态总体样本协差阵的精确分布》。20 世纪 40 年代,多元统计分析被应用于心理学、生物学和教育学等诸多领域。随着计算机的诞生,多元统计分析在社会学、医学等领域得到进一步的应用。然而,直到上世纪七八十年代,多元统计分析在我国才逐渐得到重视[1]。因此,相比统计学和数学的其他方向,多元统计学在我国仍为一门相对较新的学科。

随着大数据时代的来临,多元统计分析理论和方法的重要性日益凸显,目前已成为机器学习、模式识别、数据挖掘等数据学科和方向的专业基础课程。据调查,在我国统计学专业开设的专业课中,《多元统计分析》已成为第二大受欢迎的课程,近八成修过该课程的学生认为该课程很有帮助,受欢迎程度仅次于《统计计算与软件》[2]。为顺应大数据时代发展潮流,加快实现将上海海事大学建设成为世界一流的海事大学的总体规划目标,满足我校数学、统计学、物流和航海等相关专业的发展需求,学校先后于 2015 年和 2016 年面向应用统计硕士生和数学系本科生开设多元统计分析(必修)课程,至今已成功开课 7 次。同时,学校还通过数学建模、上海市大学生科创项目等竞赛或实践活动开设多元统计方法及软件应用系列辅导讲座,在全校范围内普及多元统计分析技术,进一步拓展了本门课的知识内涵和外延。

2. 航海类院校多元统计分析的教学现状及存在问题

目前,多元统计分析课程已成为航海类高校数学、统计学及其他数据分析相关专业的一门核心课程,具有较强的应用性。下面以上海海事大学为例介绍该课程的教学现状及存在的问题。

在知识层面上,多元统计分析可视为一元统计学的推广。因其涉及多变量,需借助高等代数(线性代数)中的向量和矩阵语言来简化表述方式。同时,还需用到数理统计中的期望、协方差矩阵和假设检验等相关概念。因此,学好本门课程需要学生具备较好的高等代数(线性代数)和数理统计基础。我校数学系本科生的《高等代数》和《概率论与数理统计》分别安排在大一的上下两学期和大二下学期开课,而《多元统计分析》自 2018 年 9 月起安排在大三上学期开课(原安排为大三下学期),这造成部分学生遗忘了学过的高等代数和部分数理统计知识。比如,在学习《多元统计分析》中的主成分分析时,部分学生忘记如何计算矩阵的特征值及标准正交特征向量。为解决此问题,任课教师在讲解本门课时,先用约 3 学时的时间系统复习矩阵论和数理统计知识点,同时督促学生课后及时复习,跟上课程节奏。

在教材内容处理方面,传统教材并未明确地将课程核心内容进行凝练。为帮助学生对比学习,笔者在讲课时将课程核心内容分为三个模块:基础知识(矩阵、数理统计、多元分布及统计推断)、分类方法(判别分析和聚类分析)以及降维方法(主成分分析和因子分析)。部分学生经过前两年的数学专业学习,已经习惯于“概念 + 做题”的学习模式,这在一定程度上忽视了数学(统计)思想,缺乏数学和统计学的宏观视野。同其他数学学科一样,多元统计分析需要用到较多的数学工具。不同的是,多元统计分析是一门应用性极强的学科,学生只有借助软件解决现实问题后才能感受到其魅力。因此,传统的“概念 + 做题”的学习模式不适合本门课程。这也是对教材内容进行整合的出发点。通过这样的内容重新编排,可以进一步优化知识结构,启发学生探索同类方法的异同点,提升统计鉴赏能力。

在教学方式方面,笔者采用“理论+实验+文献阅读”的方式开展教学。客观上讲,多元统计分析课程中的符号复杂、数学推导繁琐,学生在初学时容易产生畏难心理,在一定程度上会影响其学习积极性。为及时了解学生的真实学习情况,笔者于每次开课学期开学一个月后作匿名问卷调查,根据调查结果及时调整教学方式。表1为2017~2018下学期课程开设一个月后对我校2015级49名数学专业学生的问卷反馈结果。结果显示,16.33%的学生在之前四周学习随机向量、多元分布等相对抽象和复杂的知识点时,学习积极性较低。通过对问卷的其他结果进一步分析表明,反映课程“很难”的同学多存在基础不扎实(挂科)或没及时复习已遗忘的矩阵和数理统计基础知识的问题,导致学习包袱越背越重。众所周知,激发学生的学习积极性和主动性是教学得以顺利开展的必要条件。为提高学生学习兴趣,考虑到本门课的实用特点,在后期实际教学过程中采用“理论(板书/课件)+实验”为主、文献阅读为辅的教学模式,平均每6学时的理论课程配以3学时的实验课和适量的(课堂或课后)文献阅读课。在具体讲授判别分析、聚类分析、主成分分析和因子分析等内容时,采用形象教学和启发式教学,充分调动学生的学习兴趣,鼓励其将学到的统计方法用于解决数学建模、科创项目和毕业论文中遇到的问题。

Table 1. Feedback on *multivariate statistical analysis* (Grade 2015)

表 1. 学生对《多元统计分析》课程的学习反馈(2015级)

难易度	人数	比例
容易	1	2.04%
中等	40	81.63%
很难	8	16.33%

3. 航海类院校多元统计分析课程教学的几点思考

多元统计分析揭示随机变量之间的相互关系以及内在变化规律,具有社会科学和自然科学的双重特征。目前,本门课程在我校数学、应用统计等专业的开设以及全校范围内的数学建模、大学生科创、毕业论文等方面的广泛应用,具有较大的课程影响力。结合我校航海特色,笔者阐述在讲授多元统计分析课程时所做的相关教学探索:

3.1. 了解学科历史,激发学习兴趣

要想学好一门课,必须充分了解其发展历史以及相关概念(原理)产生的背景。因此,笔者特别重视本门课程的发展历史,在教学时尽可能向学生讲述该学科核心理论的来龙去脉。比如,在讲授判别分析中的鸢尾花案例时,介绍该案例的由来及其与现代统计学奠基人之一 Fisher 的联系;在讲解主成分分析时,介绍其提出人、统计学家 Pearson 的相关故事,并在课堂上指导学生阅读该方法的推广人、统计学家 Hotelling 的原始文献[3]中的部分内容;在讲解因子分析时,介绍 Pearson 和英国心理学家 Spearman 在 20 世纪初定义智力所作的统计分析工作。

由于历史原因,西方国家的统计学发展早于我国,所以多元统计分析课程中常出现的统计学家亦多为西方学者,如 Wishart, Fisher, Hotelling, Roy 等。在课程初期介绍多元统计发展史时,学生往往会产生这样的疑问:为什么多元统计学发展历史上见到的似乎都是西方学者,却很难看到中国学者的身影?为让学生了解我国老一辈统计学家在多元统计分析学方面的工作,笔者以多元统计分析开创人之一、我国著名统计学家许宝騄先生为例,介绍许先生在多元统计分析领域做出的奠基性工作。当学生通过斯坦福大学 T.W. Anderson 教授纪念许先生的文字[4]了解到许先生在上个世纪三四十年代便已从事多元统计

分析数学理论前沿工作并且培养了一大批中外统计学家时，所有学生都感到极大的惊喜和敬佩，为老一辈统计学家的开创精神深受鼓舞，民族自豪感油然而生。

3.2. 实施案例教学，注重能力培养

由于我校是一所以航运、物流、海洋等实践性较强的学科为特色的院校，相当一部分学生毕业后会从事与之相关的工作。因此，学校格外重视培养学生解决实际问题的能力。因此，从学校学科融合发展的角度来讲，客观上也需要数学、统计学等相关学科作为学校特色学科的知识后盾，并在数学、统计学等课程的教学过程中体现学校的办学特色。然而，传统的数理统计课程教法多侧重理论，疏于学生的实践和实际应用能力的培养。学生学完之后只知理论，遇到实际问题无从下手。

为提高学生解决现实问题的意识，本课程采用“理论 + 实验”的交叉教学模式，充分利用案例教学法阐释多元统计理论，其中理论课时与实验课时之比为一般为 2:1。课程采用的案例覆盖航海、通讯、经济、金融、体育、生物等学科领域。此外，笔者注重搜集学校特色专业中衍生的实际科研案例。比如，在讲解主成分分析时，笔者指导学生搜集船舶轨迹追踪相关数据并用所得数据进行主成分分析，初步尝试解决船舶航行路线缺失数据补全和预测问题。在这项实践活动过程中，很多同学深刻体会到统计数据的采集不易，也切身感受到统计学的魅力，反映以后拟继续深造学习从事航海数据处理相应工作。此时，笔者结合中国首位穿越北冰洋的女驾驶员、我校白响恩老师以及我校其他两位参加“雪龙号”南极科考老师的先进事迹，鼓励学生勇于科研探索，爱校荣校，为我国航海事业做贡献。

3.3. 利用形象教学，力求化繁为简

如前所述，多元统计分析课程因符号复杂、推导繁琐等客观问题，造成学生在初学时极易产生畏惧心理。这就要求教师在授课时做到形象教学，在学生初学时将统计思想从复杂的数学符号中剥丝抽茧出来。在前期的教学过程中，笔者采用“理论课 + 实验课”的交叉授课模式。然而，在授课过程中，发现一部分学生忽略统计理论学习，认为多元统计分析理论和原理过于繁琐乏味，不如案例有趣。这种现象反映了这部分学生未能从繁琐的公式中掌握统计思想。因此，如何将公式和原理化繁为简，用最直观的方式、以学生喜闻乐见的形式讲述多元统计原理是本课程教学中亟需解决的问题。

为帮助学生掌握统计思想，在介绍每个重要的多元统计方法前，先引入学生感兴趣的图片、音乐、视频、生活实例或相应的统计学家轶事，通过类比或历史还原讲清该方法的本质，然后再引导学生用精准的数学语言描述。经过“引入 - 类比 - 重述”的过程，学生实际上已经在老师的启发下几乎自主地推导出相应的统计分析方法，充分调动了学生的学习积极性。比如，在讲解主成分分析时，笔者没有采用传统的数学推导方式直接引入，而是先设问：怎样拍照更好看？几乎所有学生的兴趣一下子被该问题调动起来了，纷纷给出自己的答案。由教师或其他同学归纳并总结出：不要正对着或侧对着脸拍，可以从斜上方按一定角度拍摄。在此基础上，进一步设问：为什么这个角度拍出照片好看些？学生通过积极讨论和教师的引导，逐渐发现高维向低维空间投影时信息会有损失，必须找到最好的投影方向才能尽可能保持原貌。在此基础上，教师引出给高维数据“拍照”（即数学上的投影），借助矩阵特征值和线性相关等线性代数知识逐渐引出主成分分析。图 1 以给水壶拍照为例阐释数据降维（投影）和主成分分析的思想，其中左图用来鼓励学生选出最能辨别为拍摄物为水壶的方向（红线），右图为沿着左图中的红线轴垂直旋转直至找到第二长轴（绿线）。启发并解释做出这样旋转的原因，并告知水壶上的红线和绿线分别对应于水壶的“第一主成分”和“第二主成分”。通过这样的教学设计，学生学习的积极性被充分调动起来，更便于其把握统计思想，欣赏统计之美。学校教务系统教学反馈显示，学生十分喜欢这种教学方式，认为该教法更易于把握统计思想。

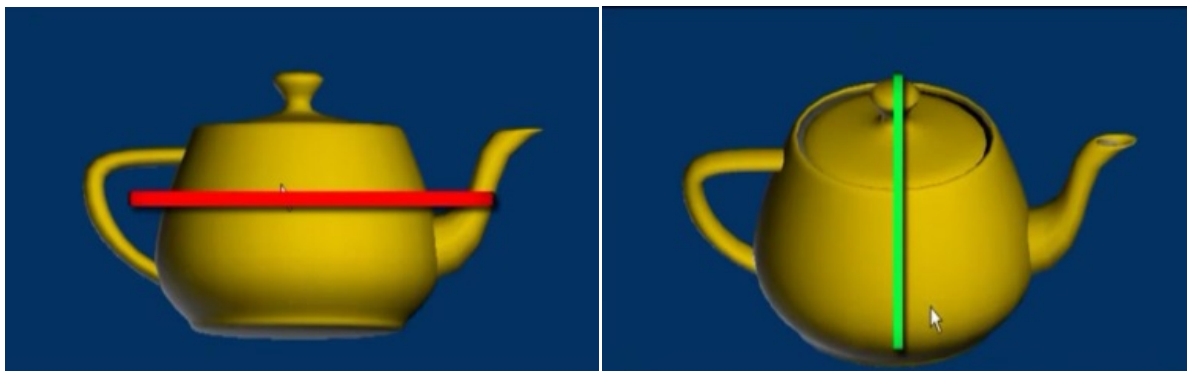


Figure 1. Exemplifying the idea of projection and principal component analysis by taking pictures of a kettle
图 1. 以给水壶拍照为例阐述投影思想及主成分分析技术

3.4. 开设翻转课堂，鼓励自主学习

近年来，作为数据科学的基础课程，多元统计分析课程内涵随着数据科学的发展变得越来越充实。因此，学生只有通过不断学习，才能从总体上把握多元统计分析的最新进展。然而，部分学生缺乏自主学习的动力和能力，造成课程培养目标和现实的脱节。为鼓励学生自主学习，自本门课 2015 年首次开课以来，笔者便不断尝试在开课班级开展翻转课堂(文献阅读)活动，即每学期用 3 学时左右的课堂课时及相应的课外时间让学生提前阅读与课程相关的学术文献。引导学生根据阅读提示，在完成文献阅读后试讲或撰写阅读报告。在之后的实验课上，邀请学生用翻转课堂期间搜集的数据进行演算，并给出统计解释。对学生的试讲、阅读报告和实验结果，教师给予点评。此项活动已在 4 届数学系本科生中开展，使学生通过自主学习掌握相应的多元统计知识，提高运用统计知识解决问题的能力，熟悉科技论文的写作方法，为毕业论文的撰写做好铺垫。

4. 小结

本文以航海类院校的多元统计分析课程为例，阐述了如何利用学校特色和课程特点开展多种形式的教学活动，激发学生学习兴趣，培养学生统计素养，提高其利用统计技术自主解决现实问题的综合能力。

基金项目

本文工作得到国家自然科学基金项目(11601323)和上海海事大学文理学院重点课程建设项目(2018)资助。

参考文献

- [1] 韩明. 多元统计分析: 从数据到结论[M]. 上海: 上海财经大学出版社, 2016: 1.
- [2] 教育部统计学类专业教学指导委员会. 我国统计学类专业本科教育现状的调查与分析[J]. 统计研究, 2015, 32(2): 104-108.
- [3] Hotelling, H. (1933) Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, 24, 417-441. <https://doi.org/10.1037/h0071325>
- [4] Anderson, T.W. 许宝騄在多元分析方面的工作[J]. 数学的实践与认识, 1980(3): 8-12.

知网检索的两种方式：

1. 打开知网首页：<http://cnki.net/>，点击页面中“外文资源总库 CNKI SCHOLAR”，跳转至：<http://scholar.cnki.net/new>，搜索框内直接输入文章标题，即可查询；
或点击“高级检索”，下拉列表框选择：[ISSN]，输入期刊 ISSN：2331-799X，即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版：<http://www.cnki.net/old/>，左侧选择“国际文献总库”进入，搜索框直接输入文章标题，即可查询。

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：ces@hanspub.org