

Teaching Practice and Research of Probability and Statistics under the Background of Big Data

Tingting Zhang Lili Fan Jiemei Zhao Qingfang Wang Xiaowu Zhou

School of Mathematics and Computer Science, Wuhan University of Light Industry, Wuhan, Hubei, 430048, China

Abstract

Probability and Statistics is one of the most important mathematics core courses in higher education institutions. As a public basic course for science, engineering, and management majors, it uses the basic theory and methods of random phenomena to lay a foundation for subsequent professional learning. The ultimate goal of this course is to cultivate students' ability to shift from deterministic thinking to stochastic thinking, and to enhance their comprehensive skills in collecting, organizing, and analyzing data to make statistical inferences about practical problems and make reasonable decisions. Probability and statistics, through the deep integration of theory and practice, has become a pivotal discipline in higher education that connects mathematical foundations with cutting-edge applications. Its core value lies in endowing students with modern scientific literacy to "think about the world with data and control the unknown with probability".

Keywords

Probability statistics; Integration of theory and practice; Modeling

大数据背景下《概率统计》教学实践与研究

张婷婷 范丽丽 赵杰梅 王庆芳 周小伍

武汉轻工大学数学与计算机学院, 中国·湖北 武汉 430048

摘要

《概率统计》是高等学校的最重要的数学公共核心课之一。作为理工、经管类专业的公共基础课,利用随机现象的基本理论与方法,为后续专业学习打下基础。本课程应将培养学生从确定性思维转向随机性思维,提升学生采集、整理、分析数据对实际问题进行统计推断并制定合理决策的综合能力作为最终目标。概率统计通过理论与实践的深度融合,成为高等教育中连接数学基础与前沿应用的枢纽学科,其核心价值在于赋予学生“用数据思考世界,以概率驾驭未知”的现代科学素养。

关键词

概率统计; 理论实践融合; 建模

1 引言

大数据时代的到来,给概率统计这门学科带来了机遇和挑战。《概率统计》作为高校数学基础课中最重要的课程之一,其应用几乎遍及自然科学和社会科学的各个领域,尤其在经济管理、工程技术、人工智能、医学科学等交叉领域具有非常重要的应用价值。如何在大数据背景下采取一种高效、新颖的教学方法进行《概率统计》的教学,已成为当前教育研究的热点问题^[1-5]。

随着计算机的运算速度和存储能力的发展,收集、存储数据越来越简单。掌握的数据量越来越大,数据存在形式的复杂度越来越高。如何在如此庞大的数据中得到有用的信

息,这就是数据挖掘理论要解决的问题。针对这一问题,需要运用数据分类分析、聚类分析及时间序列分析等统计学方法来进行处理。数据科学的理论基础是统计学,数据科学可以看作是统计学在研究范围和分析方法上不断扩展的结果。利用统计方法来解决大数据背景下的各类问题,具有非常重要的意义。统计与建模是相互依存的。统计学为建模提供数据分析和推断的理论基础,而建模则是将统计方法转化为解决实际问题的结构化工具。高校通过数学建模竞赛(如全国大学生统计建模大赛,中国计算机挑战赛-大数据挑战赛)训练学生将统计理论转化为解决问题的能力。统计学与计算机科学、经济学等学科交叉,培养“数据建模+领域知识”的复合型人才,这将是一个提升教学质量、培养创新人才的有效途径,也将是推动教育改革的重要举措。本文的研究目标是围绕实践应用来开展教学活动,淡化理论教学。缩短概率论的理论教学时间,增加统计方法和应用的课时比例。最后通过参加建模比赛,使学生的综合能力得到进一步的提

【作者简介】张婷婷(1983-),女,中国湖南邵阳人,博士,副教授,从事数据分析与处理、偏微分方程解的存在性研究。

升。该教学模式能有效促进知识转化与能力发展，为培养新工科背景下的复合型统计人才提供了可推广的解决方案。

综上所述，本文探索大数据背景下《概率统计》课程教学模式与方法具有重要的理论价值和现实意义。统计学和数据科学相结合的交叉学科兼具理论创新价值与实践应用潜力。在实践环节有效提升了学生的数学建模及跨学科解决问题能力，为高校创新人才培养提供了新的实施路径。

2 调整教学内容比例

传统教学方法一直是重概率，轻统计。概率主导课程，导致教学内容失衡。这些问题的存在有着多方面的原因：一方面，多数高校因学时限制（如48学时），仅能讲授概率论、参数估计、假设检验部分，统计核心内容（回归分析、时间序列）被迫转为自学。过度强调古典概型排列组合技巧、积分计算等手工解法，忽视数据标准化、统计量计算等实践技能。另一方面，教师对概率论内容更熟练，统计应用能力薄弱，导致教学重心天然倾斜。传统教学主要依赖手工计算，没有充分应用统计软件（如SPSS、Matlab、Python）处理实际大规模数据，学生实际应用能力没有得到很好地培养。因此，遇到数据处理与分析问题时束手无策，只能临场学习统计方法，现学现用。统计应用能力不足，导致毕业论文选题狭窄，研究深度受限。商科、医学等专业学生缺乏统计工具应用能力，影响后续课程（如计量经济学）学习^[1-2]。

在概率统计教学过程中淡化理论，注重实践应用，压缩概率论的授课时间比例，增加统计方法和应用的课时比例。修订教学大纲，调整教学计划，在有限的教学时间内，加强实践性教学内容。通过建模大赛让学生对理论知识有更为直观的认识，深刻体会如何应用概率统计里的方法来解决实际的问题。激发学生的学习兴趣，提高学生的创新能力。

因此，突破“重概率轻统计”的传统教学方式，推动概率统计从数学演绎向数据科学实践转型，其核心在于以实际问题为基点，以统计工具为桥梁，培养学生“用数据预测及决策”的科学素养。

3 理论联系实际 —— 数据驱动下的统计建模

概率统计是一门应用性非常强的学科。可是很多学生学完概率统计，不知其到底如何使用。上课听得懂，课后会做题，但不会应用。主要原因有以下两点：（1）没真正理解概率统计的本质意义；（2）缺乏必要的应用训练。所学知识进行应用是学习的目的，同时也是学习的动力。教学中理论联系实际，让学生感受到概率统计强有力的应用性，能激发学生的学习兴趣 and 探索精神，开阔他们的视野，培养学生的应用能力和创新能力。本文强调概率统计的实际应用，将统计方法与数据分析处理进行深度融合，通过统计方法进行数据分析与处理，从而对实际问题作出预测或者决策。问题解决过程核心框架如下：

3.1 数据处理

统计数据是进行统计分析的前提。数据处理是将收集到的各种原始数据条理化、系统化，使之符合统计分析的要求。通过数据处理可以简化数据，更有效地提供统计信息。在数据处理之前首先应该进行数据预处理。数据预处理主要包括数据清洗和特征工程：数据清洗是通过异常值检测、缺失值填补提升数据质量，为后续分析奠定基础。特征工程是利用相关性分析、主成分分析筛选关键变量，降低维度冗余。

3.2 模型构建与验证

统计模型是以概率论为基础，通过数学方程描述变量间随机关系，其核心在于揭示数据内在规律并进行推断决策。统计模型主要包括回归模型、聚类分析、单指标模型、变系数模型、时间序列模型、随机森林模型、方差分析模型、主成分分析模型等等。利用最小二乘法估计参数，残差分析检验拟合优度。

3.3 预测和决策

统计模型预测与决策是通过数学框架将数据分析结果转化为可执行策略的过程。在对模型进行选择与验证后，利用模型对相应问题给出预测和决策对现实具有重要的价值。

下面我们就回归分析的建模思路，进行具体的介绍。首先给出回归分析建模思路示意图如图1所示：

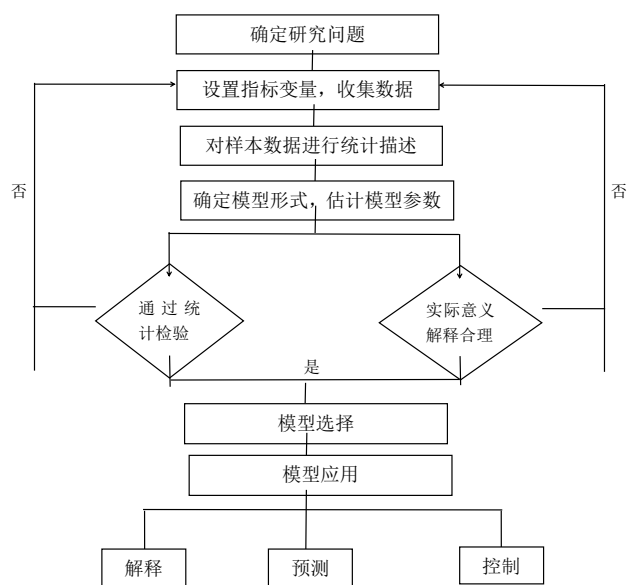


图1 回归分析建模思路示意图

具体步骤如下：

- （1）在建立模型之前，确定具体的研究问题。
- （2）根据研究问题确定指标变量的范围。如果总体得不到，可以通过样本采集方式收集数据。
- （3）了解数据的统计特征，如观察样本数据的分布状态，通过变量间的散点图探索变量间是否存在线性关系。通过了解数据的统计特征，可以迅速掌握数据的特点和问题，可以初步判断回归模型的适用性。

(4) 确定回归模型的形式并利用样本数据对模型中的参数进行估计。常用的参数估计方法有最小二乘法、最大似然法。

(5) 进行统计检验和模型解释。如果统计检验的结果不能满足分析要求,则需要回到第二步,重新考虑纳入模型的变量指标是否合理。模型建立的目的是解决实际问题,因此模型解释合理是评价模型效果的重要考量。如果参数结果的解释与理论相背离,此时也应该回到第二步,考虑变量指标是否需要调整。最后需要得到既满足统计检验要求又解释合理的模型。

(6) 在备选模型中选择最优者。在构建理论模型时,并没有确定模型的具体形式。需要在众多备选模型中根据某种评价准则选择一个最优模型。模型选择和模型评价是融合在一起的。

(7) 将回归模型应用于实际具体问题。

在教学过程中培养学生数据处理、模型构建、预测和决策的能力尤其重要。只要掌握的统计方法,将以上过程实践到具体问题,对于其他类似问题也能够迎刃而解。回归分析是统计里最重要的一类模型,在社会经济与自然科学领域均应用广泛,在教学过程中应该对此内容的重点讲解,培养学生数学建模的能力以及创新能力。因此,教学中理论联系实际,既能激发学生的学习兴趣 and 探索精神,又能让他们真正理解知识,并使应用能力、创新能力得到提高。

4 培养学生统计软件的应用能力

解决实际问题时,数据可能非常庞大,笔算是不现实的,统计软件能帮助解决计算上的问题。因此,学生在学习过程中应掌握统计软件的使用。教师授课时结合统计软件演示例题的计算,适当讲授统计软件的使用方法,学生能力将得到极大地提高。

统计软件主要包括 Stata、SPSS、SYSTAT、NCSS 2025、Matlab、Python、R 语言。Stata 主要进行高级面板数据分析,适合混合线性模型,在经济学,流行病学领域有着广泛的应用。它具有命令行高效操作,实时更新分析模块的特点。SPSS 主要用于交互式方差分析、聚类分析、机器学习集成,在社会科学应用较多。它具有菜单式操作,AI 增强可视化的特点。Python 是大数据分析中广泛使用的统计软件,需要学生熟练掌握 Python,并能应用于各种经济数据分析中。

统计软件繁杂多样,在较短的课时安排里不可能实现一一讲解。因此,在教学过程中,可以着重讲解 Python 在统计学中的具体应用,其他软件提供学习资料帮助学生在有需要的时候进行自学。

5 建模大赛提升学生解决问题的综合能力

建模竞赛考察非常广泛,学生需要掌握多方面的理论知识,并对它进行综合应用。这种多维度训练机制能显著提

升学生解决实际问题的综合能力。主要培养的能力有以下几个方面:

5.1 问题拆解与转化能力

面对复杂现实问题,参赛者需将需求转化为可量化数学模型,训练抽象化问题本质的能力。

5.2 跨学科知识整合

融合数学理论、算法设计及领域知识,培养多维度解决方案设计能力。赛题常涉及环境治理、智能制造等前沿领域,需调用交叉学科知识库。

5.3 数据处理能力

数据预处理主要包括异常值识别,缺失。通过数据处理可以简化数据,更有

效地提供统计信息。常用的方法有时间序列分析及关联规则挖掘。

5.4 算法应用能力

将机器学习中的算法应用到建模里。常用的算法有 CNN 图像识别、LSTM 时序预测、Apriori 关联分析。

5.5 统计软件应用能力

掌握 Python 等统计分析软件。

5.6 团队协作能力

建模大赛一般是组队参加,团队如何分工合作,如何相互配合能达到最好的配合效果。

5.7 提出创新方案

在已有的研究基础之上,加入创新点,使得问题的解决的更好。

6 结语

总之,《概率统计》课程的教学实践如何得到更好的应用迫在眉睫,本文通过教学内容课时比列的改变、加强理论与实践相结合,参加建模大赛等举措这一新的教学模式来实施教学过程。这一模式将会提升学生的学习兴趣,真正体会学以致用乐趣。新的教学模式的实施也对学生提出了更高的要求,更有利于培养学生的自学能力和解决实际问题的能力,从而使学生在今后的学习和工作实践中具备更强的竞争能力。

参考文献

- [1] 阿布力克木·阿布都热依木. 高校概率统计课程教学方法的研究[J]. 南昌教育学院学报,2012,27(9):69-70.
- [2] 姚敏. 大学概率统计教学体系改革模式探讨[J]. 吉林省教育学院学报,2013,29(8):149-150.
- [3] 李建军,刘力维. 面向能力培养的教学模式探讨——以工科概率统计课程为例[J]. 大学数学,2014,30(1):107-110.
- [4] 侯学刚. 概率统计教学模式与方法改革研究[J]. 教育教法探讨与实践,2014, 12:57.
- [5] 丁芳清,汪忠志,方益. 工程教育认证背景下工科院校概率论与数理统计课程教学改革与实践[J]. 安徽工业大学学报(社会科学版),2024,41(06):51-53.