

基于R软件“概率论与数理统计”课程的教学改革初探

李 静¹, 李雪艳², 魏传华³

¹中国劳动关系学院应用技术学院, 北京

²中国民航大学理学院, 天津

³中央民族大学理学院, 北京

收稿日期: 2022年11月22日; 录用日期: 2022年12月21日; 发布日期: 2022年12月29日

摘 要

本文主要介绍R软件在“概率论与数理统计”课程教学中的使用, 用以提高学生的学习兴趣, 使学生对该课程的知识有更深刻的认识, 最终提高学生的数据分析能力。

关键词

概率论与数理统计, R软件, 数据分析

Research on the Teaching Reform of the Course “Probability Theory and Mathematical Statistics” Based on R Software

Jing Li¹, Xueyan Li², Chuanhua Wei³

¹School of Applied Technology, China University of Labor Relations, Beijing

²School of Science, Civil Aviation University of China, Tianjin

³School of Science, Minzu University of China, Beijing

Received: Nov. 22nd, 2022; accepted: Dec. 21st, 2022; published: Dec. 29th, 2022

Abstract

This paper mainly discusses the use of R software in the teaching of the course “Probability and

Mathematical Statistics”, which is used to improve students’ interest in learning, enable students to have a deeper understanding of the knowledge of the course, and ultimately improve students’ data analysis ability.

Keywords

Probability and Mathematical Statistics, R Software, Data Analysis

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

随着科学技术的发展, 我们已经进入了人工智能与大数据时代, 正处于新一轮的产业变革和科技革命的浪潮中。在这种形势下, 我国本科人才的培养受到高度重视。数学在人才培养中一直具有重要的不可替代的作用。而对于非数学专业的学生来说, 数学知识的学习主要通过“微积分”、“线性代数”和“概率论与数理统计”这三类课程。由于专业的不同, 这三类课程的学分学时和教学内容各有不同。不同于其他两个课程, “概率论与数理统计”是一门专门讲授随机数学的课程, 培养学生的随机性思维。

由于课程学时学分的设置, 教材的选取以及教师本身教学方法等多方面的问题, 目前“概率论与数理统计”教学中在一些问题[1], 比如“满堂灌”的教学模式, 陈旧的教学案例, 所讲知识与实际问题脱节等。这些问题造成的后果是学生对该课程的学习兴趣不高, 对知识点理解不深, 无法利用所学知识分析和解决实际问题, 最终造成学生数学基础不牢固, 实践创新能力较低, 从而达不到人才培养目标。

本文主要针对上述问题, 提出将 R 软件引入到“概率论与数理统计”的教学中, 在不需要占用过多学时的情况下, 基于该软件改变教学内容和教学模式, 除了提高学生的兴趣之外, 更能培养学生的数据分析能力, 从而提高学生的学习效率, 提升学生实践创新能力。虽然目前已经有文献[2] [3] [4]也研究了概率论与数理统计课程中使用 R 软件, 但都是基于个别案例进行介绍, 没有系统性说明选用 R 软件的必要性等。

2. R 软件介绍及选用理由

R 软件是一个有着强大统计分析功能和作图功能的软件系统, 是由奥克兰大学统计学系的 Robert Gentleman 和 Ross Ihaka 研发。目前能够处理数据和统计学问题的软件非常多, 比如常见的统计软件 SAS、SPSS、Stata、R, 还有随着大数据和人工智能发展起来的 Python, 以及数学软件 matlab 和 Mathematics 等。

在“概率论与数理统计”课程中选用 R 软件, 主要有如下几个理由。

第一, R 软件免费, 对所有学生来说不存在使用条件上的限制。除了 R 和 Python 之外, 上面提到的软件都是商业化软件, 有的价格几万元, 有的一年的使用费就高达十几万元, 因此不是所有学校都有条件购买和使用这些软件。相比而言, R 软件完全免费, 所有资源都在其官方网站 <https://www.r-project.org> 上。

第二, R 软件的下载和安装非常方便, 适合课堂上即安即用, 特别适合统计软件的初学者, 尤其是那些对软件使用有畏惧情绪的学生。有的统计软件安装文件大, 安装起来耗时较长。有的软件比如 Python 安装方式多样, 给初学者带来困难。R 软件的下载途径单一, 且安装文件本身很小, 以 2022 年 11 月 R 官网提供的供 Windows 系统使用的 4.2.2 版本为例, 安装文件大小仅为 75.9 MB, 从下载到安装只需要三

五分钟即可完成,教师和学生在课堂上完全可以即安即用。

第三,与 SPSS 和 Stata 等软件相比, R 不是局限性地下拉菜单式操作模式,而是一种灵活的编程语言,有强大的统计计算和绘图功能,能够自主实现一些具体的需求,这与“概率论与数理统计”课程教学改革的需求一致。对于初学者来说,虽然语言操作模式比菜单操作模式听起来困难,但 R 是一种解释型语言,并不是编译语言如 C 语言对程序的结构要求较高。比如某班 20 个同学的数学成绩,对于这 20 个数据可以很简单的定义为 $x \leftarrow c(86, 75, \dots, 92)$, 输入 $\text{mean}(x)$ 就是求取均值, $\text{var}(x)$ 就是求取方差, $\text{hist}(x)$ 和 $\text{boxplot}(x)$ 就是分别画这组数据的直方图和箱线图。此外 R 软件的使用可以培养学生的编程思维,提高学生的编程能力,这是菜单式软件所不具备的。

第四, R 软件的使用为学生后续统计方法和量化方法的学习起到了良好的开端。不论是理工科还是人文社会学科,专业领域的数据分析和量化建模都非常重要。不管学生本科毕业后走向工作岗位,还是继续攻读研究生深造,都需要进一步学习数据分析和量化方法,特别是在目前的大数据和人工智能时代。“概率论与数理统计”课程中 R 软件的引入为后续的学习起到了重要的引导作用。

3. 基于 R 软件的教学改革设计

“概率论与数理统计”教学中引入 R 软件,能够提高学生的学习兴趣,作为辅助手段能够帮助学生理解一些知识点的本质,从而更好的掌握知识。此外还加强了学生数据分析能力的培养和提高,这对学生的后续学习以及解决问题至关重要。下面给出几个具体的教学改革案例。

现在使用的《概率论与数理统计》教材,在概率分布的介绍上,一般都是直接给出数学公式,将知识点的讲解作为纯数学知识的介绍,学生的学习重点为数学公式的记忆和用公式计算相关问题,很少引导学生去思考概率分布和实际数据之间的关系。以泊松分布为例,一般的教材都是直接给出该分布的定义,并简单提及具有泊松分布的随机现象在实际生活中很多,比如一本书一页中的印刷错误数、某地区在一天内邮递遗失的信件数等等。对于该知识点的学习,应该不仅仅让学生记住具体的分布公式,更要引导学生理解这个理论分布和生活中的实际数据之间的关系,使学生能够深刻认识到用概率分布去刻画现实生活中呈现出的一些规律的必要性,从而也体会到数学抽样能力的强大。因此,为了让学生对这个分布有直观感受,应该先给学生展示生活中真实的数据,让学生自己去思考这些数据呈现出的规律。但由于客观条件限制,教师不可能花费较多的时间和精力去收集大量的真实数据,这种情况下可以采用 R 软件能够产生指定分布的随机数的功能,去产生大量的随机数,供学生观察,基于灵活的 R 软件结合图形让学生观察大量的数据,从而让学生深刻了解理论分布与实际数据之间的关系。

对于连续型分布里最为常见的正态分布,该分布的介绍应该先让学生看实际数据,然后基于直方图观察数据的分布规律,然后从对直方图的讨论自然而然的引入概率密度函数曲线。通过这个过程,让学生深刻的认识到正态分布的本质,从而也自然而然的引出统计推断任务,如果这个数据集呈现出的规律是一个正态分布规律,那么该用哪一个具体的正态分布去刻画它,即用均值和方差取值为多少的正态分布去刻画这组实际数据,这就引出了参数估计的问题,为后面参数估计的学习进行了前期介绍。

大数定律和中心极限定理知识在“概率论与数理统计”课程中属于难点,现有教材也是侧重数学证明和直接使用,不能让学生对定理有深刻的认识。在这部分的教学,可以基于 R 软件采用数值模拟的方式,让学生直接动手操作感受样本均值是如何随着样本量的增大逼近总体均值,不同的分布其和的形式如何向正态分布逼近。

4. 总结

大数据和人工智能时代对当代大学生的培养提出了更高的要求,如何提高学生的学习兴趣,如何培

养和提高学生的实践创新能力，如何提高学生的综合素质，这都是当今高等教育面临的问题。在重要数学课程“概率论与数理统计”中引入 R 软件，对教学内容和教学方式进行改革，能够提升学生的学习兴趣，能够让学生对所学知识有更为深刻的认识。

基金项目

中国劳动关系学院教育教学改革立项项目(JG1406); 中央民族大学本科教学改革创新项目(CX2007)。

参考文献

- [1] 吴新军, 郭朕, 潘冬. 概率论与数理统计课程教学改革探索与研究[J]. 大学教育, 2021(10): 125-127.
- [2] 张志成. R 软件在概率论与数理统计课程教学中的应用[J]. 河南工学院学报, 2020(5): 78-80.
- [3] 安丽霞, 卢丑丽, 燕扬. 初探 R 软件在独立院校概率论与数理统计教学中的应用[J]. 大学教育, 2018(9): 102-104.
- [4] 沈新娣, 侯丽英. 医学类专业概率论与数理统计教学实践: R 软件与案例结合教学[J]. 教育教学论坛, 2020(22): 284-285.