

基于 MATLAB 的数理统计教学难点 可视化教学研究

刘勇波^{1,2,3} 楚鑫^{1,2,3}

1. 内蒙古师范大学 数学科学学院, 呼和浩特;
2. 无穷维哈密顿系统及其算法应用教育部重点实验室, 呼和浩特;
3. 内蒙古自治区应用数学中心, 呼和浩特

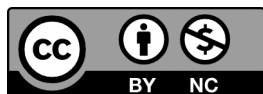
摘要 | 数理统计课程作为统计学专业的一门专业必修课, 在构建学生系统的统计理论框架、培养数据分析方法掌握能力及培养应用型人才等方面都起着重要作用。传统数理统计教学常因概念抽象、逻辑复杂而面临诸多难点。针对学生对抽象定理和概念理解困难的痛点, 本文以样本均值的分布、假设检验的 p 值为例, 探讨利用MATLAB强大的计算与图形可视化功能, 将抽象定理与复杂计算转化为直观的动态图形, 以深化学生对核心统计思想的理解。

关键词 | 数理统计; MATLAB; 样本均值; 假设检验

Copyright © 2025 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

<https://creativecommons.org/licenses/by-nc/4.0/>



1 引言

数理统计课程是统计学及众多相关专业的核心必修课, 其重要性源于在知识构建、能力培养和目标达成三个方面不可替代的战略价值。它不仅是知识体系中的重要一环, 更是承前启后的关键枢纽, 是培养科学思维和解决实际问题能力的重要平台。通过本课程的学习, 学生应掌握数理统计的基本知识和原理, 理解数据分析的基本思想和方法, 具备严谨的数据处理和统计分析能力, 为从事科学研究、解决实际问题及统计建模奠定坚实基础。然而, 在数理统计的传统教学范式下, 其教学过程往往被繁琐的公式演绎、抽象的定理证明与大量复杂的数值计算所主导。这种侧重于纯理论推导的教学模式, 极易导致课堂氛围枯燥乏味, 致使部分学生产生畏难情绪, 进而逐渐丧失学习兴趣与主动性。更深层次的问题在于, 由于教学过程中直观性与交互性的普遍缺失, 学生对于许多核心统计思想的理解往往停留在“知

其然不知其所以然”的层面。例如, 在讲授“样本均值的抽样分布趋近于正态分布”这一中心极限定理时, 学生由于无法目睹随着样本量增大, 分布形态动态收敛于钟形曲线的过程, 仅能被动接受结论, 因而对其合理性与普适性心存疑虑。又如, 在阐释假设检验中的“ p 值”这一核心概念时, 传统的讲授方式难以将其“在原假设成立的前提下, 观察到当前样本数据或更极端情况的概率”这一定义, 转化为一个清晰可视的尾部概率面积。学生只能机械记忆“ $p < 0.05$ 则拒绝原假设”的准则, 而无法从统计推断的本质理解 p 值作为衡量证据强度尺度的真正意义, 最终陷入“概念混淆、应用僵化”的困境。

在此背景下, 引入高效、直观的计算工具辅助教学, 已成为推动数理统计教学改革的重要路径。MATLAB作为全球范围内应用最为广泛的高性能科学计算软件之一, 凭借其一系列核心优势, 在数理统计的教学与学习等领域中, 是一款不可多得强大辅助工具^[1-3]。首先,

基金项目: 内蒙古自治区一流学科科研专项 (YLKZX-NSD-008)。

通讯作者: 楚鑫, 内蒙古师范大学数学科学学院, 讲师, 博士, 研究方向: 应用数学。

文章引用: 刘勇波, 楚鑫. 基于MATLAB的数理统计教学难点可视化教学研究 [J]. 教育研讨, 2025, 7 (10): 1068-1074.

<https://doi.org/10.35534/es.0710201>

MATLAB的优势体现为工程化的简洁性,它以矩阵为基本数据单元,语法风格高度接近于数学公式的标准书写形式,这极大地降低了学生的编程门槛。其次,MATLAB具有计算的高效性与环境的集成性,能够快速处理大规模数值计算。最重要的是,MATLAB具有强大的可视化功能,能够轻松地将抽象的理论与枯燥的数据转化为生动的二维、三维乃至动态图形。无论是绘制概率分布曲线、展示中心极限定理的收敛过程,还是可视化假设检验,都能将数理统计中“只可意会”的核心思想变为“清晰可见”的直观影像,从根本上深化学生的理解,是传统“黑板+PPT”教学模式无法比拟的卓越辅助^[4-6]。

本文拟结合具体教学案例,如样本均值的分布、假设检验的 p 值,系统探讨MATLAB在数理统计教学中的应用,并分析其在提升教学效果和深化学生理解能力的作用,从而推动数理统计课程教学的多样化发展。

2 MATLAB 在数理统计教学中的具体应用案例

为解决数理统计中概念和定理抽象难懂的问题,下面将展示MATLAB如何将抽象的数学定理转化为直观的视觉呈现,从而有效提高教学效果,帮助学生更深入地理解知识本质。

2.1 样本均值的分布

定理1: 设 $x_1, x_2, \dots, x_n$ 是来自某个总体的样本, $\bar{x}$ 为样本均值。

(1) 若总体分布为正态分布 $N(\mu, \sigma^2)$, 则 $\bar{x}$ 的精确分布为 $N(\mu, \sigma^2/n)$;

(2) 若总体分布未知或不是正态分布, 其期望为 μ , 方差为 σ^2 且存在, 则当 n 较大时 $\bar{x}$ 的渐进分布为 $N(\mu, \sigma^2/n)$ ^[7]。

该定理指出, 无论总体服从何种分布, 只要样本容量足够大, 样本均值就近似服从正态分布。为验证此结论, 本研究利用MATLAB进行了如下实验:

首先, 自正态总体 $N(1, 4)$ 中随机生成60万个数据, 将其分为2万组, 每组30个数据。使用以下代码实现:

```
n = 30;           % 样本容量
total_groups = 20000; % 组数
mu = 1;           % 总体均值
sigma = 2;         % 总体标准差
total_samples = total_groups * n;
data = mu + sigma * randn(total_samples, 1);
data_reshaped = reshape(data, n, total_groups);
```

其次, 对每一组样本的30个数据求其样本均值。代码如下:

```
sample_means = mean(data_reshaped);
```

最后, 利用MATLAB中的histogram 函数画出全部数据的直方图, 如图1所示, 样本均值的直方图如图2所示。此外, 为了与正态分布进行对比, 本研究还用normpdf函数分别画出了总体和样本的理论正态分布的概率密度图。

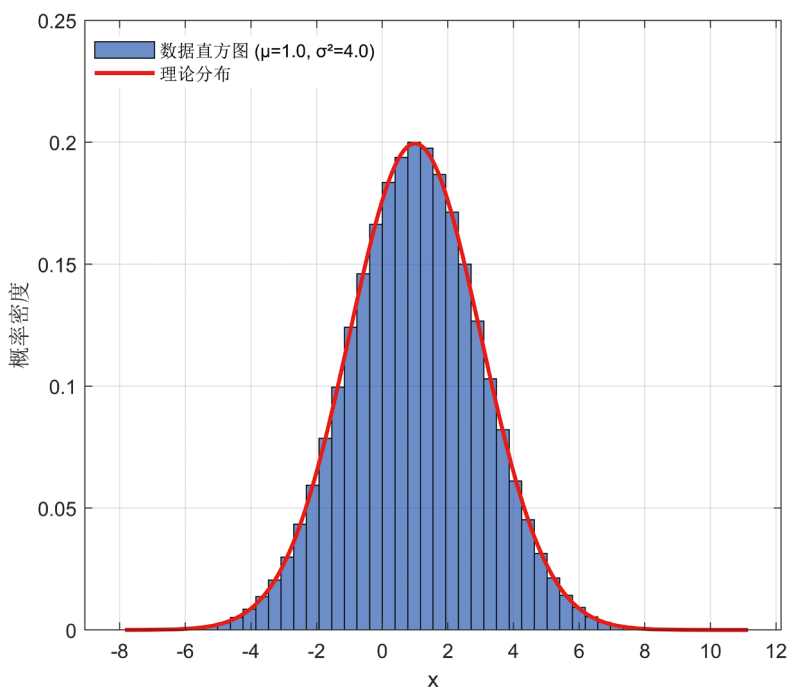


图 1 总体分布 $N(1, 4)$

Figure 1 The population distribution $N(1, 4)$

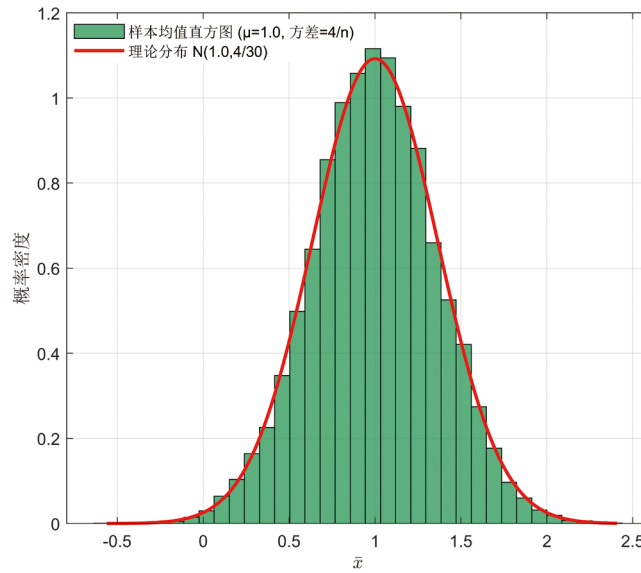
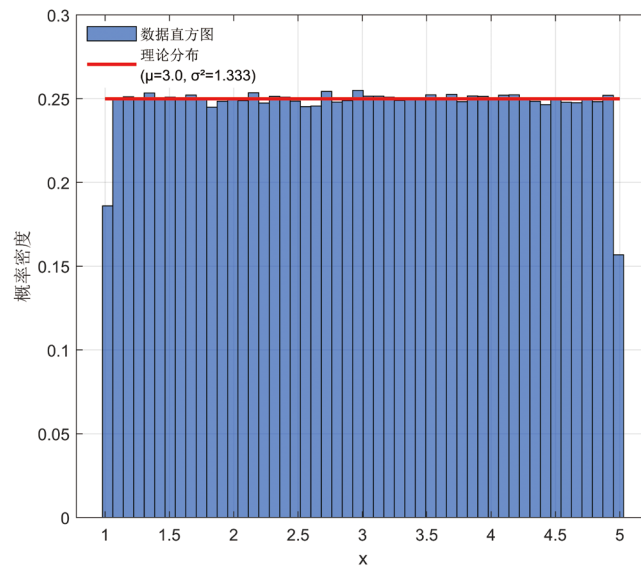
图 2 样本均值分布 ($N=30$)Figure 2 The sampling distribution of the sample mean ($N=30$)

图1绘制了全部原始数据的直方图及其理论概率密度曲线，图2展示了2万个样本均值的直方图及其理论正态分布密度曲线。结果表明，当总体分布为正态分布时，其样本均值也服从一个正态分布，这验证了定理1中的第(1)条结论。

为验证定理1的第(2)条结论，使用MATLAB的rand函数从均匀分布 $U(1, 5)$ 中生成60万个数据，并分成2万组，具体代码如下：

```
n = 30;           % 样本容量
total_groups = 20000; % 组数
```

```
total_samples = total_groups * n;
a = 1;          % 均匀分布下限
b = 5;          % 均匀分布上限
data = a + (b - a) * rand(total_samples, 1); % 从均匀分布
抽样
data_resaped = reshape(data, n, total_groups);
随后，本研究针对均匀分布总体 $U(1, 5)$ 重复了
图1、图2中的实验（如图3、图4所示），结果支持了定
理1中的第(2)条结论，表明即便总体为非正态分布，
其样本均值也收敛于正态分布。
```

图 3 总体分布 $U(1, 5)$ Figure 3 The population distribution $U(1, 5)$

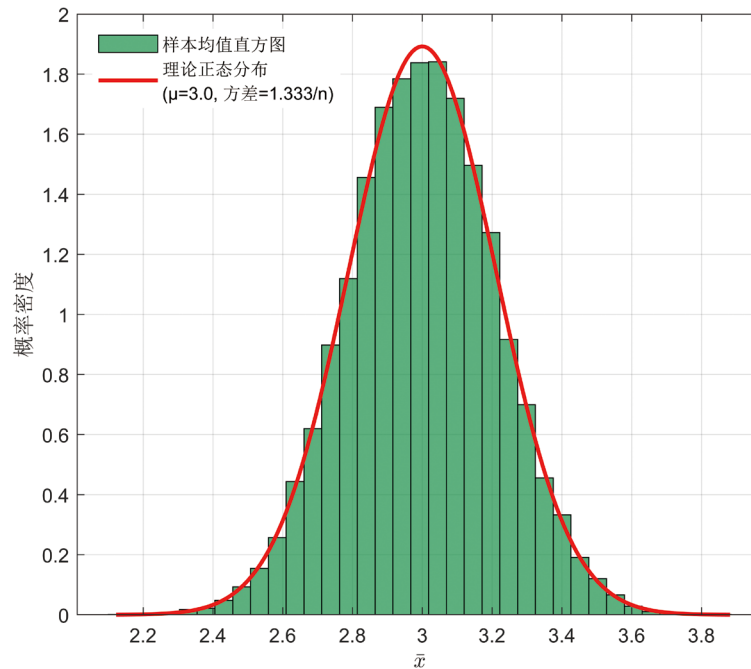


图 4 样本均值分布 ($N=30$)

Figure 4 The sampling distribution of the sample mean ($N=30$)

2.2 假设检验的 p 值

定义1: 在一个假设检验问题中, 利用样本观测值能够作出拒绝原假设的最小显著水平称为检验的 p 值。

p 值的决策规则为: 如果 $p < \alpha$, 则在显著性水平 α 下拒绝原假设 H_0 ; 反之, 如果 $p > \alpha$, 则在显著性水平 α 下接受原假设 H_0 。实际中, p 值很小时 (如 $p \leq 0.001$) 即可拒绝原假设, p 值很大时 (如 $p > 0.5$) 即可接受原假设, 只有当 p 值与 α 接近时才需要比较^[7]。

尽管 p 值是假设检验的核心, 但其概念非常抽象。学生常常产生困惑, 如“显著性水平 α 究竟在检验统计量概率密度图中的哪个位置?” “ p 值很小, 为什么就能拒绝原假设?”, 为此, 本研究借助MATLAB进行可视化展示, 以阐明 p 值的真实含义。

在一个假设检验问题中, p 值是在原假设成立的前提下, 获得当前样本观测结果或更极端结果的概率。如图5所示, p 值具体表现为从观测 t 统计量 (黑色实线) 向右延伸的尾部区域面积。当 p 值很小 (如 $p=0.035$) 时, 表明若原假设成立, 则当前样本结果出现的概率极低 (仅3.5%)。由于 p 值小于显著性水平 $\alpha=0.05$, 而观测 t 统计量落入了拒绝域 (红色虚线以右的区域), 这表明小概率事件在一次抽样中实际发生, 从概率角度分析是不太可能的, 从而使得本研究有理由怀疑原假设的真实性。因此, p 值越小, 反对原假设的证据就越强, 该图通过 p 值区域与拒绝域的位置关系, 直观解释了假设检验的决策逻辑。

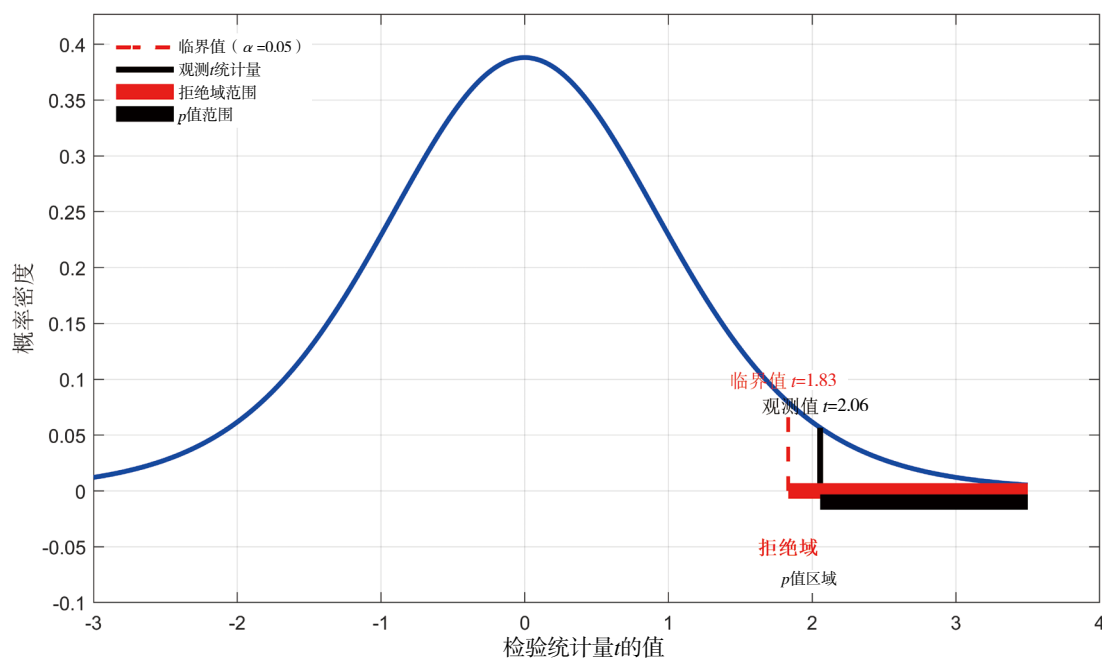
图5的MATLAB代码如下:

```
clear; clc;
%% 设置检验参数
mu0 = 100; % 原假设的总体均值
alpha = 0.05; % 显著性水平
% 样本数据
sample_data = [102.5, 103.2, 101.8, 104.1, 102.9, 103.5,
101.2, 104.8, 102.1, 103.9];
%% 计算统计量
n = length(sample_data); % 样本容量
x_bar = mean(sample_data); % 样本均值
s = std(sample_data); % 样本标准差
% 计算观测到的t检验统计量
t_observed = (x_bar - mu0) / (s / sqrt(n));
% 计算p值 (右侧检验)
df = n - 1; % 自由度
p_value = 0.035; % 直接设定p值为0.035
t_observed = tinv(1 - p_value, df); % 根据p值反推t统计量
figure('Position', [100, 100, 900, 500]);
set(gcf, 'Color', 'w');
% 生成t分布曲线
x = linspace(-3, 3.5, 1000);
y = tpdf(x, df);
% 绘制t分布曲线
plot(x, y, 'b-', 'LineWidth', 2.5);
```

```

hold on
xlabel('检验统计量 t 的值', 'FontSize', 12);
ylabel('概率密度', 'FontSize', 12);
grid on;
% 计算临界值
t_critical = tinv(1 - alpha, df);
% 绘制临界值线（红色虚线）
critical_line = line([t_critical, t_critical], [0, tpdf(t_critical,
df)], ...
'Color', 'r', 'LineStyle', '--', 'LineWidth', 2);
% 绘制观测到的统计量（黑色实线）
observed_line = line([t_observed, t_observed], [0, tpdf(t_
observed, df)], ...
'Color', 'k', 'LineStyle', '-', 'LineWidth', 3);
% 标记拒绝域范围（只在x轴上标记）
reject_region = plot([t_critical, max(x)], [0, 0], 'r-',
'LineWidth', 8);
% 标记p值范围（只在x轴上标记）
pvalue_region = plot([t_observed, max(x)], [-0.01, -0.01],
'k-', 'LineWidth', 8);

```



注：自由度 $df=9$, $p=0.035$, $\alpha=0.05$ 。

图 5 假设检验 p 值的可视化演示

Figure 5 Visual demonstration of the p -value in hypothesis testing

3 MATLAB 在数理统计教学应用中的优势和劣势

在数理统计教学的辅助软件选择中，与Python和R软件相比，MATLAB有独特的优势与局限。其具体优劣势如下。

3.1 MATLAB 在数理统计教学应用中的优势

第一，MATLAB (Matrix Laboratory) 的基石是矩阵，而数理统计中的许多计算（如回归、主成分分析）本质都是矩阵运算。其底层对线性代数运算进行了极致优化，执行效率非常高。

第二，语法相对简单，特别适合有数学背景的用户快速入门，降低了教学和学习的初始门槛。

第三，MATLAB提供了极其丰富和高质量的可视化函数（如plot、scatter、histogram、surf等），能够轻松绘制2D、3D图形，用于探索性数据分析、分布验证和结果呈现。

第四，在教学中，学生可以轻松地将统计公式（如协方差矩阵 X^*X ）直接翻译为直观的MATLAB代码，有助于理解计算本质。

第五，利用其高效的随机数生成器（如rand、randn）和向量化计算，可以轻松实现大规模的蒙特卡洛模拟，用于验证统计定理（如中心极限定理）、评估统计量的抽样分布、计算Bootstrap置信区间等。

第六，Histogram Tool等交互式工具，允许学生动态调整参数并立即看到效果，极大地加深了对统计概念

(如箱数对分布形态的影响)的理解。

3.2 MATLAB 在数理统计教学应用中的劣势

第一, MATLAB是商业软件, 个人版和校园版许可价格昂贵。这限制了其在个人学习者或预算有限的学校中的普及。相比之下, R和Python是完全免费的。

第二, 对于复杂、不规则的数据整理、清洗和重塑任务, MATLAB的表格(table)类型功能虽然不断进步, 但在灵活性和生态丰富性上, 仍不及R的tidyverse或Python的pandas生态系统, 数据“预处理”体验相对繁琐。

第三, 其生态系统主要由MathWorks公司主导, 社区贡献的第三方工具箱在数量、多样性和活跃度上, 远不及R的CRAN和Python的PyPI。许多最新、小众的或特定领域的统计方法可能只在R或Python中提供。

4 教学实施建议

前文探讨了MATLAB在解决数理统计教学难点中的具体应用, 并客观分析了其优劣。然而, 要将可视化教学的潜力转化为切实的教学成果, 还需制定周密的实施策略。同时, 随着技术的发展, 也应展望其未来的融合方向。

4.1 分层递进的教学实施建议

为确保MATLAB有效服务于教学, 而非增加师生负担, 建议采用分层递进的实施路径。

第一, 教师主导演示层(初级阶段): 在教学初期, 应以教师操作演示为主。教师精心准备可视化案例, 在课堂上动态展示关键概念(如中心极限定理的模拟)。此阶段的目标是“化抽象为具体”, 激发学生兴趣, 降低入门畏难情绪。课件和演示脚本应由教研室统一规划、开发, 确保其科学性与教学适配性。

第二, 学生模仿验证层(中级阶段): 当学生具备一定基础后, 可提供“代码模板”或“函数库”, 引导学生开展模仿性、验证性实验。例如, 教师提供绘制不同分布曲线的框架代码, 学生通过修改参数观察分布形态的变化; 或提供假设检验的流程代码, 学生代入自有数据验证理论结果。这一阶段旨在巩固理论知识, 培养学生的编程动手能力。

第三, 项目探究应用层(高级阶段): 在课程中后期, 鼓励学生以小组形式开展小型探索性数据分析项目。学生需自主完成从问题定义、数据搜集(或由教师提供)、MATLAB程序编写, 到结果分析与报告撰写的全过程。这不仅让学生综合运用所学统计知识, 更能培养他们解决实际问题的能力、团队协作精神和科学素养。教师的角色也将从知识传授者转变为项目指导者和促进者。

4.2 与课程考核方式的融合

教学方法的改革必须辅以考核方式的革新。应突破

“一张试卷定乾坤”的传统模式, 将MATLAB应用能力纳入形成性评价体系。例如, 平时作业中可设置需编程完成的数据分析题; 课程中期安排小项目汇报; 期末考核亦可引入开卷上机考试, 考查学生利用MATLAB解决新问题的能力。这种多元化评价机制, 能更全面地反映学生的综合能力, 同时引导学生的学习重心从“死记公式”转向“理解与应用”。

5 结语

数理统计课程中的核心概念与定理往往抽象难懂, 构成了教学过程中的主要难点。为破解这一困境, 本研究通过具体教学实例表明, 借助MATLAB强大的计算与可视化功能, 能够将抽象的数学定义与定理(如样本均值的分布规律、 p 值的本质)转化为直观的图形呈现, 从而有效化解学生的认知障碍, 提升教学成效。将MATLAB融入教学过程, 不仅是一种创新的教学手段, 更是推动课程从理论灌输向能力培养转型的关键举措。

在大数据时代与产出导向教育理念的驱动下, 熟练掌握MATLAB等科学计算工具变得愈发重要。未来, 需持续深化教学改革, 将此类工具深度嵌入课程体系, 使学生能够切实掌握利用技术解决实际问题的能力。

综上所述, 对于侧重统计思想直观理解、计算实验与可视化的数理统计教学而言, MATLAB是一个极具价值的平台, 它能将抽象的统计概念生动具象化, 为学生应对未来的学术与职业挑战奠定坚实基础。

参考文献

- [1] 田智. 推荐一个优秀的科学与工程计算编程语言—MATLAB[J]. 集宁师专学报, 2003, 25(4): 11-16.
- [2] 张清叶, 尚邵阳. 基于MATLAB的随机性数学思维培养的探索与实践[J]. 科技资讯, 2022, 20(23): 140-143.
- [3] 高发玲, 姚中华, 孙建英. 引入MATLAB实验的概率论与数理统计线上线下混合教学的研究与实践[J]. 甘肃科技, 2022, 38(9): 66-68, 72.
- [4] 敬成林, 李宇. MATLAB在数理统计假设检验中的应用[J]. 科技资讯, 2017, 15(17): 214, 216.
- [5] 胡红娟, 杜健, 鞠桂玲, 等. 融入数学建模思想和MATLAB的中心极限定理形象化教学[J]. 信息系统工程, 2020(8): 167-168, 170.
- [6] 杜世强, 刘华. MATLAB在概率论与数理统计课程教学过程中的应用[J]. 甘肃科技, 2019, 35(17): 42-43.
- [7] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程[M]. 北京: 高等教育出版社, 2019.

Research on Visualization Teaching of Teaching Difficulties in Mathematical Statistics Based on MATLAB

Liu Yongbo^{1,2,3} Chu Xin^{1,2,3}

1. College of Mathematics Science, Inner Mongolia Normal University, Hohhot;

2. Laboratory of Infinite-dimensional Hamiltonian System and Its Algorithm Application, Hohhot;

3. Center for Applied Mathematical Science, Inner Mongolia, Hohhot

Abstract: As a compulsory professional course for statistics majors, the course of mathematical statistics plays a significant role in helping students build a systematic theoretical framework of statistics, cultivating their ability to master data analysis methods, and fostering applied talents. Traditional teaching of mathematical statistics often encounters many difficulties due to its abstract concepts and complex logic. To address the problem point that students have difficulty understanding abstract theorems and concepts, this paper takes the distribution of sample means and the p -value in hypothesis testing as examples to explore the use of MATLAB's powerful computing and graphic visualization functions, transforming abstract theorems and complex calculations into intuitive dynamic graphics, so as to deepen students' understanding of core statistical ideas.

Key words: Mathematical statistics; MATLAB; Sample mean; Hypothesis testing