Creative Education Studies 创新教育研究, 2025, 13(5), 600-606
Published Online May 2025 in Hans. https://www.hanspub.org/journal/ces
https://doi.org/10.12677/ces.2025.135382

案例教学法在《统计计算》课程中的实践

刘　凯

上海立信会计金融学院统计与数学学院，上海

收稿日期：2025年3月26日；录用日期：2025年5月20日；发布日期：2025年5月29日

摘　要

在大数据时代背景下的复杂数据需要更加高级的统计方法，《统计计算》这门课程从两个方面提升统计专业学生专业能力：(1) 把理论的统计方法通过编程变成可靠、高效的算法。(2) 借助现代计算机的处理能力，发展新的统计方法。因此，这是一门对编程能力要求更高的课程。针对应用统计系专业学生编程能力薄弱，采用案例演示教学法，充分把控理论方法和实践教学的难度和深度，因材施教，使学生不仅能够应用统计方法解决实际问题，也能理解背后的理论原理。

关键词

教学改革，案例演示教学法，统计计算，统计编程思维

The Practice of Case Teaching Method in Statistical Computing Course

Kai Liu

School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai

Received: Mar. 26th, 2025; accepted: May 20th, 2025; published: May 29th, 2025

Abstract

In the era of big data, more advanced statistical methods are required to deal with complex data. "Statistical Computing" is a course to improve the professional ability of statistical students from two aspects: (1) The theoretical statistical methods are implemented into reliable, efficient computational algorithm. (2) Develop new statistical methods with the help of the processing power of modern computers. Therefore, this is a course that requires more programming ability. Since the programming ability of students majoring in Applied Statistics is weak, the case demonstration teaching method is introduced in the Statistical Computing course to balance between the theoretical and practical difficulty and depth. Thus, students will not only be able to apply statistical

文章引用: 刘凯. 案例教学法在《统计计算》课程中的实践[J]. 创新教育研究, 2025, 13(5): 600-606.
DOI: 10.12677/ces.2025.135382

methods to solve practical problems, but also understand the theoretical principles behind them.

Keywords

Teaching Reform, Case Demonstration Teaching Method, Statistical Computing, Statistical Programming Thinking

Copyright © 2025 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

随着现代科技的迅速发展，全球进入了信息革命的时代，各行各业数据收集越来越广泛。在迅速积累的海量数据中包含了许多以前无法触摸的现象和规律，对海量数据进行分析，从海量数据中发现规律，已经成为统计学和信息科学的热门研究方向。高校对统计学人才的培养也随之发生了变化。

目前大部分高校统计专业在本科课程设置中都设置了与主要专业必修课相配套的统计软件实验课程，以培养学生数据收集、数据整理和数据分析能力。基础的统计软件实验课程侧重使用现成的内置函数或程序包进行经典统计方法的实现，但并不重视经典统计方法的编程过程。另外，随着计算机的发展，发展出了借助现代计算机处理能力的新的统计方法，称为计算统计。统计计算正是将经典统计方法的编程实现和计算统计汇总在一起。

部分高校开设了《统计计算》这门课程，作为选修课供学有余力的高年级本科生修读。国内外《统计计算》课程大多采用 R 语言作为编程软件，R 语言作为最主要的开源统计软件之一一直深受高校统计学者的青睐。Python 作为一门通用语言，随着其在各行各业越来越广泛的应用，以及其在机器学习领域丰富的软件包，也使其成为应用统计学人才必备的一项技能。我校自 2018 年开设了《统计计算(Python)》以 Python 作为教学软件，供应用统计系和金融统计系本科三年级学生选修。经过 3 年教学实践，在 2020 级的培养方案中将该课程设置成专业必修课，可见该课程的重要性。这门课程内容包括(1) 基本统计计算方法的编程实现，例如，随机数的生成算法，计算分布函数、分位数函数的一般方法，统计推断可能涉及到的优化方法、近似计算方法、矩阵计算方法、随机模拟积分，(2) 由计算方法发展出来的统计方法，例如，Boostrap、MCMC 等。一些学者也对统计计算课程进行了教学改革探索[1]-[4]。本文从教学难点出发，从教材的选择、教学方法的改革方面对统计计算课程进行一些探索和实践，以期得到较好的教学效果。

2. 《统计计算(Python)》教学难点

该课程的核心目标是让学生掌握更底层的统计方法思维并建立统计编程思维，非常锻炼学生的底层代码编写能力。目前教学中主要存在以下问题和难点。

(1) 学生编程基础较弱。虽然在前期必修课程中学习了《Python 程序设计基础》《统计应用软件》等基础编程课，但在教学过程中发现学生虽然所有软件都会一点，但都不精、编程基础不扎实的问题。比如，学生虽然学过基础代码编写，但如何编写一个 Python 函数，使得代码具有普适性和重复使用性的编程思维还十分欠缺。

(2) 本校统计学专业有部分学生高中是选文科的，因此，在教学中如何把控深度，平衡文理科背景学生对理论知识的吸收并将实现编程是教学中的难点之一。

3. 教学改革思路

3.1. 教材

该课程一开始采用的高惠璇著《统计计算》[5]，该教材更偏重理论，而《统计计算》作为一门结合理论和应用的课程，并且是面向具有一定统计知识的大三学生开设的课程，更加侧重将统计方法变成可靠、高效的算法并编程实现。因此，本课程在教学改革过程中更换主要教材为李东风编著的《统计计算》[6]，该教材更加明确地阐述了积分、最优化、矩阵计算这些数学问题跟统计问题之间的关系。另外，李东风版本的《统计计算》包含了 R 代码示例，很好地平衡了理论和实践的比例，针对理科思维相对偏弱的学生，可以更加侧重实践部分的内容。另外，针对 Python 编程，选用嵩天等人编著的《Python 语言程序设计基础》作为辅助教材[7]。

3.2. 实时演示法

传统软件实验课程教学通常采用讲授式教学方法，即对事先准备好的代码进行讲解。这样的方式在代码的讲解过程中学生容易走神，抓不住代码编写过程中的难点。因此，在教学方法改革上，试采用实时演示法，即在课堂上教师通过实时编写代码进行演示，同时要求学生在教师演示过程中同步自己动手进行编写，这样更有助于学生直观了解代码的编写过程，有效地让学生注意力更加集中。

3.3. 案例教学法

案例教学法是哈佛大学在二十世纪二十年代提出的，它以分析实际案例来进行教学。哈佛的案例教学法在很多商科管理学科中使用较广泛，收到较好的教学效果。二十世纪九十年代开始，国内教育界开始探讨案例教学法[8]-[10]。统计计算课程内容较抽象，选取经典和贴合生活实际的案例进行课堂教学，能够更有助于学生理解深刻理解概念，从而实现统计方法的代码编写。在本课程教学过程中，在部分知识点中引入了案例教学法，以探索案例教学法在统计计算课程中的教学效果。

4. 案例教学法的实施

下面结合两个知识点的教学介绍案例教学法在《统计计算》课程中的应用情况。

4.1. 统计研究中的随机模拟教学的案例

引入案例：研究北京市的年轻人对下一年房价涨跌预测。

问题拆解：在这个案例中，目标变量是下一年房价的涨跌，其取值为涨或跌，可以用统计学中的二项式分布来描述该目标变量。假设 X 是二项式变量，其值取 1 表示房价上涨，取 0 则表示房价下跌。用 p 表示房价上涨的概率，则 X 服从概率为 p 的二项式分布。统计研究的主要任务之一是通过样本对总体进行推断。这里，通过计算 p 的置信区间，对房价涨跌进行预测。根据中心极限定理，可建立基于正态分布的看涨概率的区间估计。1927 年， Wilson 提出了一个修正的置信区间，称为威尔逊置信区间。为了比较两种置信区间，在没有真实数据的情况下，我们可以通过随机模拟比较这两种置信区间。好的置信区间应该满足以下两个条件：(1) 相同置信度下，置信区间越短越好；(2) 覆盖率大于等于置信度，且两者越接近越好。

试验设计：随机模拟需要进行多次重复实验，因此在代码编写过程中，需要对单次抽样估计进行重复，如何编写一个可重复使用的 Python 函数是重点。首先，教师一步步对单次试验代码进行编写计算两个置信区间。假设房价看涨概率为 0.1，样本量为 30，置信水平为 95%，设计如下单次试验的代码计算两种置信区间以及两个置信区间是否包含真实的看涨概率。

```
# 一种情况
sig = 0.95
n = 30
p = 0.1

sample = np.random.binomial(1,p,size=n)
phat = np.mean(sample)
lmd = stats.norm.ppf(1-(1-sig)/2)
pdif = phat-p
rad1 = lmd*np.sqrt(phat*(1-phat)/n)
if (abs(pdif)<rad1): cover = 1
else: cover=0
diff1 = 2*rad1
ptilde = (phat + lmd**2/(2*n))/(1+lmd**2/n)
delta = lmd/np.sqrt(n)*np.sqrt(phat*(1-phat)+lmd**2/(4*n))/(1+lmd**2/n)
if (abs(ptilde-p)<delta): cover2 = 1
else: cover2 = 0
diff2 = 2*delta
print(cover,cover2,diff1,diff2)
```

在代码开头，给定一组置信水平、样本量和看涨概率(见上文代码 2~4 行)，并随机生成二项式随机数(见上文代码第 6 行)，然后采用基本代码模块对研究内容编写，例如在 7~13 行是基于正态的置信区间，14~18 行是 Wilson 点估计和置信区间。19 行输出了两个方法的置信区间覆盖指标(即置信区间是否包含真实值，若包含记为 1，若不包含记为 0)，以及区间长度。由于随机模拟要重复多次，假设随机模拟试验重复 100 次，随后将上述代码模块作为基模块，在其基础上进行扩张，如下所示。

```
# 重复m次
sig = 0.95
n = 30
p = 0.1
N = 100

i = 0
ans = np.zeros((N,4))
for i in range(N):
    sample = np.random.binomial(1,p,size=n)
    phat = np.mean(sample)
    lmd = stats.norm.ppf(1-(1-sig)/2)
    pdif = phat-p
    rad1 = lmd*np.sqrt(phat*(1-phat)/n)
    if (abs(pdif)<rad1): cover = 1
    else: cover=0
    diff1 = 2*rad1
    ptilde = (phat + lmd**2/(2*n))/(1+lmd**2/n)
    delta = lmd/np.sqrt(n)*np.sqrt(phat*(1-phat)+lmd**2/(4*n))/(1+lmd**2/n)
    if (abs(ptilde-p)<delta): cover2 = 1
    else: cover2 = 0
    diff2 = 2*delta
    ans[i] = (cover,cover2,diff1,diff2)
np.mean(ans,axis=0)
```

从上述代码可见，在原来代码的基础上，增添了第 5 行进行 100 次重复模拟，在 7~9 行添加了控制循环的指标，10~22 行对计算过程进行重复，23 行将每一次结束输出保存在矩阵的一行，最后对所有结果取平均值。上述计算过程中，限定了二项式分布的置信度、样本大小、看涨概率以及随机模拟的次数。最后通过编写函数使得这几个指标可以随意设定，具体代码如下所示。

```
1.  def  conf(sig,n,p,N=10000):
2.          i  =  0
3.          ans  =  np.zeros((N,4))
4.          for  i  in  range(N):
5.                  sample  =  np.random.binomial(1,p,size=n)
6.                  phat  =  np.mean(sample)
7.                  lmd  =  stats.norm.ppf(1-(1-sig)/2)
8.                  pdif  =  phat-p
9.                  rad1  =  lmd*np.sqrt(phat*(1-phat)/n)
10.                 if  (abs(pdif)<rad1):  cover  =  1
11.                 else:  cover=0
12.                 diff1  =  2*rad1
13.                 ptilde  =  (phat  +  lmd**2/(2*n))/(1+lmd**2/n)
14.                 delta  =  lmd/np.sqrt(n)*np.sqrt(phat*(1-
    phat)+lmd**2/(4*n))/(1+lmd**2/n)
15.                 if  (abs(ptilde-p)<delta):  cover2  =  1
16.                 else:  cover2  =  0
17.                 diff2  =  2*delta
18.                 ans[i]  =  (cover,cover2,diff1,diff2)
19.         result=ans.mean(axis=0)
20.         return(result)
```

上述代码中删除了置信度、样本大小和看涨概率的指定值，对随机模拟的次数给定了一个默认值，因为该指标在随机模拟中通常是固定的。这里注意，前面我们使用 100 次随机模拟，这里的默认值是 10000 次。原因在于，基础代码编写过程中，通常用一个比较小的随机模拟次数进行代码调试，在基代码调试完成后，在正式的使用代码中使用较大的数，保证随机模拟的可靠性。最终，选取不同的置信水平、样本量和看涨概率，计算不同情况下两种置信区间的覆盖率，以研究两种置信区间计算方法的表现。代码如下所示。

```
21.     df = np.array([])
22.     for k in [0.99,0.95,0.90,0.8]:
23.         for size in [30,120,480]:
24.             if (size==30): prbs=[0.1,0.3,0.5,0.7,0.9]
25.             if (size==120): prbs = [0.05,0.1,0.3,0.5,0.7,0.9,0.95]
26.             if (size==480): prbs =[0.01,0.05,0.1,0.3,0.5,0.7,0.9,0.95,0.99]
27.             for prb in prbs:
28.                 ans = np.append([k,size,prb],conf(sig=k,n=size,p=prb))
29.                 df = np.concatenate((df,ans),axis=0)
30.
31.     ddf = pd.DataFrame(df.reshape((84,7)),
32.                    columns = ('confidence level','n','prob','coverage 1','cover-
age 2','diff1','diff2'))
33.
34.     print(ddf)
```

结果显示，即使样本量达到 480，基于正态分布的置信区间其覆盖率低于置信水平，特别当真实的看涨概率 p 靠近 0 或 1 的情况下，效果更差。Wilson 置信区间的覆盖率与置信度基本一致。

4.2. Boostrap 的案例

引入案例：研究某地小学五年级学生身高和体重的关系。

问题拆解：要研究身高和体重的关系，可以用相关系数这个统计量来量化，所以这个问题就转化为通过样本数据推断总体身高体重的相关系数。这里需要计算相关系数的估计值以及置信区间，并与 0 比较。若置信区间包含 0，则说明没有足够的证据证明体重跟身高相关，若置信区间不包含 0，则说明体重跟身高相关。

试验设计：首先，输入 10 个样本的身高体重数据，并计算样本相关系数，代码如下所示。

```
35.    import numpy as np
36.    import matplotlib.pyplot as plt
37.
38.    h = np.array([144,166,163,143,152,169,130,159,160,175])
39.    w = np.array([38,44,41,35,38,51,23,51,46,51])
40.
41.    corr_matrix = np.corrcoef(h,w)
42.    print('协方差矩阵为',corr_matrix)
43.    corr = corr_matrix[0,1]
44.    print('相关系数的估计值为',corr)
```

其次，Boostrap 方法通过有放回地从样本中抽取 B 个样本，计算每个样本的相关系数，代码如下所示。

```
45.    sample = range(10)
46.    n = 10
47.    np.random.seed(10)
48.    nboot = 1000
49.    corb = np.zeros(nboot)
50.    for i in range(nboot):
51.        bsample = np.random.choice(sample,n,replace=True)
52.        hb = h[bsample]
53.        wb = w[bsample]
54.        corbm = np.corrcoef(hb,wb)
55.        corb[i] = corbm[0,1]
56.    # print('bootstrap 样本的相关系数为',corb)
57.    plt.hist(corb,bins=20)
58.    plt.axvline(x=corr,color='red')
59.    print(np.mean(corb))
60.    print('非参数 bootstrap 方法得到的相关系数的标准误差估计为',np.std(corb))
61.    print('非参数 bootstrap 方法得到的相关系数的置信区间下限为',np.quan-
tile(corb,0.025))
62.    print('非参数 bootstrap 方法得到的相关系数的置信区间上限为',np.quan-
tile(corb,0.9755))
```

5. 教学效果

为评估案例教学法的教学效果，选取了对应案例教学法和传统讲授法的两个课后实验报告作业的成

绩进行对比，结果如表 1 所示。案例教学法分数整体比传统教授法高，配对 t 检验结果也说明了案例教学法的教学效果显著优于传统讲授法。

Table 1. Comparison of after-class homework scores under different teaching methods
表 1. 不同教学法课后作业成绩对比

	最低分	平均分	最高分	标准差
案例教学法	80	90	95	5.25
传统讲授法	50	65	100	11.87

6. 总结

在大数据时代，各行各业对数据分析人才具有非常大的需求。作为应用统计学专业的学生，除了掌握统计方法的理论知识以外，需要熟练运用各种统计软件进行数据分析。伴随着数据的多源性和复杂性，简单的统计方法已经无法完全满足行业的需求，而经常需要在经典方法的基础上进行创新及改进，因此，在统计学人才培养过程中，培养学生使用统计方法解决实际问题，并通过编程实现算法，是统计学本科教学的核心目标之一。统计计算作为现在统计重要的一部分，是本科教学中至关重要的一门课程。编程能力的培养是依靠大量编写代码进行积累的，面对学生在编程实验课程中自主学习积极性欠缺、编程基础薄弱的问题，通过课堂上的现场代码演示，演示过程对编程逻辑的讲解，以及要求学生同步练习，引导学生从无到有地编写自己的计算函数。

基金项目

2021 年度上海高校青年教师培养资助计划。

参考文献

[1] 孙佳楠. 提升本科生“统计计算”课程学习的自我效能感的教学探索[J]. 大学数学, 2020, 36(2): 59-65.

[2] 黄性芳. 统计计算研讨课实践教学初探[J]. 科教导刊, 2017(11X): 84-85, 116.

[3] 李俊杰, 郑艺容. “统计计算”课程教学实践与探讨[J]. 科教文汇, 2019(21): 70-72.

[4] 吕书龙, 刘文丽, 梁飞豹, 薛美玉, 简彩仁. 基于发展性评价与专业素质培养的统计计算课程的教学改革研究[J]. 高等理科教育, 2020(5): 87-92.

[5] 高惠璇. 统计计算[M]. 北京: 北京大学出版社, 1995.

[6] 李东风. 统计计算[M]. 北京: 高等教育出版社, 2017.

[7] 嵩天, 礼欣, 黄天羽. Python 语言程序设计基础[M]. 第 2 版. 北京: 高等教育出版社, 2017.

[8] 沈鹏, 温立书. 案例教学法在概率论与数理统计中的应用与实践[J]. 创新教育研究, 2023, 11(11): 3608-3614. https://doi.org/10.12677/CES.2023.1111528

[9] 刘亮, 程敏, 涂文婕, 胡亚慧. 案例教学法在《大学计算机基础》课程教学中的应用探究[J]. 教育进展, 2022, 12(12): 5760-5765. https://doi.org/10.12677/AE.2022.1212875

[10] 鞠桂玲, 宋爱斌, 陈平. 案例教学法在线性代数课程教学中的应用[J]. 教育进展, 2020, 10(6): 1035-1039. https://doi.org/10.12677/AE.2020.106173