

基于偏正态分布的视角：优化高校课程成绩评定方法的探讨

田卫忠^{1*}, 韩佳仪², 张蕴初², 彭 雷¹, 丁伟康³

¹深圳技术大学大数据与互联网学院, 广东 深圳

²深圳技术大学新材料与新能源学院, 广东 深圳

³山东大学物理学院, 山东 济南

收稿日期: 2025年3月3日; 录用日期: 2025年3月22日; 发布日期: 2025年4月3日

摘 要

高校课程成绩评定传统上依赖于正态分布模型, 但在处理高分段或低分段的偏斜数据时存在局限性。本文基于偏正态分布, 探讨其在学生成绩分析中的优化应用。通过实证分析深圳技术大学四门课程(程序设计基础A、高等数学A、离散数学、面向对象程序设计)的成绩数据, 对比正态分布与偏正态分布的拟合效果。结果表明, 偏正态分布通过引入偏度参数, 能够更精准地捕捉成绩数据的非对称特性, 其拟合优度(AIC与BIC值)显著优于正态分布, 且置信区间更精确。研究进一步提出改进建议: 区分合格性考试与选拔性考试、综合过程性评价与终结性考核、增强评分透明度, 并合理利用大数据技术优化教学评估。本文为高校成绩评定提供了新视角, 支持教育评价体系的科学化与多元化发展, 助力创新人才培养目标的实现。

关键词

正态分布, 偏正态分布, 拟合精度, 成绩分析, 置信区间

Perspective Based on Skewed Normal Distribution: Exploring the Optimization of Course Grade Evaluation Methods in Higher Education

Weizhong Tian^{1*}, Jiayi Han², Yunchu Zhang², Lei Peng¹, Weikang Ding³

¹College of Big Data and Internet, Shenzhen Technology University, Shenzhen Guangdong

²College of New Materials and New Energies, Shenzhen Technology University, Shenzhen Guangdong

*通讯作者。

文章引用: 田卫忠, 韩佳仪, 张蕴初, 彭雷, 丁伟康. 基于偏正态分布的视角: 优化高校课程成绩评定方法的探讨[J]. 统计学与应用, 2025, 14(4): 76-91. DOI: 10.12677/sa.2025.144091

³School of Physics, Shandong University, Jinan Shandong

Received: Mar. 3rd, 2025; accepted: Mar. 22nd, 2025; published: Apr. 3rd, 2025

Abstract

Traditional assessment of university course grades has relied on the normal distribution model, yet this approach shows limitations when handling skewed data in high or low score segments. This study explores the optimized application of skew-normal distribution in student performance analysis. Through empirical analysis of grade data from four courses at Shenzhen Technology University (Fundamentals of Programming A, Advanced Mathematics A, Discrete Mathematics, and Object-Oriented Programming), we compared the fitting effectiveness between normal distribution and skew-normal distribution. Results demonstrate that the skew-normal distribution, by incorporating a skewness parameter, more accurately captures the asymmetric characteristics of grade data, with significantly better goodness-of-fit (AIC and BIC values) than normal distribution and more precise confidence intervals. The study further proposes improvement strategies: distinguishing qualifying examinations from selective assessments, integrating formative and summative evaluations, enhancing grading transparency, and appropriately utilizing big data technology to optimize teaching evaluation. This research provides new perspectives for academic performance assessment, supporting the scientific and diversified development of educational evaluation systems, thereby facilitating the cultivation of innovative talents.

Keywords

Normal Distribution, Skew-Normal Distribution, Fitting Accuracy, Grade Analysis, Confidence Interval

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

在教育统计学领域，学生成绩数据的分析一直是一个重要的研究课题。传统上，正态分布因其数学上的简洁性和在自然界中的广泛存在性常被用作描述学生成绩分布的基本模型。然而，随着教育数据的日益丰富和复杂化，研究者们逐渐发现许多实际情况下的成绩数据并不完全遵循正态分布，特别是成绩在低分或高分段往往表现出明显的偏斜性[1]，这种偏斜性可能由多种因素引起，如教学内容的难度、学生的学习态度、考试形式的变化等。传统的正态分布模型在处理学生成绩数据时产生了明显的局限性。

为了更准确地描述和分析这种具有偏斜特性的成绩数据，本文尝试探索其它统计分布模型。偏正态分布作为一种能够调整分布偏斜程度的扩展正态分布，因其数学表达上的灵活性和实际应用中的有效性逐渐受到关注。偏正态分布通过引入偏度参数来控制分布的偏斜方向和程度，从而能够更精细地刻画成绩数据的实际分布情况。

本文旨在探讨偏正态分布模型在大学生成绩数据分析中的应用与拟合效果。通过实证分析高校课程的学生成绩数据，比较偏正态分布与标准正态分布在拟合成绩数据方面的差异，并评估偏正态分布模型的拟合优度。本文的研究不仅有助于深入理解学生成绩数据的内在特性，还为教学评估、成绩预测及个

性化学习推荐等领域提供了新的统计方法和分析方式。

2. 高校使用正态分布评价学生成绩的由来

在高等教育体系中，课程考试成绩是教学评估的关键要素。它不仅直接体现学生对课程知识的掌握程度，还间接反映了教学质量、教学方法的有效性及学习环境的支持度。

对课程考试成绩的管理，成为了提升教学质量、优化教学评价流程的重要基石。这一过程涵盖了从考试设计、实施、评分到结果分析的全链条管理，要求教育者既要确保考试的公平性、公正性和科学性，又要能够高效地收集、整理和分析考试成绩数据，以数据为驱动，洞察教学过程中的优势与不足。长期以来，教育者认为正态分布是自然界和社会中最常见、应用最广的一种分布，并广泛应用于学生成绩评定中[2]。

当在大量随机变量上重复很多次实验时，它们的分布总和将非常接近正态分布。大量研究表明，凡符合教学规律的考试，其总体成绩应服从或接近服从正态分布，成绩特别好的和特别差的分数所占的比例小，处于中间状态的分数所占的比例大[3]。正态分布的设置可以合理体现出学生之间的区分度，因此教师在评估高校学生的成绩时，通常使学生的成绩尽可能满足正态分布。使用正态分布评估学生成绩的优势如下：

第一，影响成绩分布的因素具有随机性。学生的学习能力、努力程度和学习环境等多种因素都具有一定的随机性。这些随机因素相互作用，使得学生的成绩呈现出一种随机的分布状态。当这种随机性足够大时，成绩分布就趋向于正态分布。

第二，教育系统具有标准化特点。在高校教学中通常包含标准化的教学内容、教学方法和评价标准。这些标准化措施有助于减少个体差异对成绩分布的影响，使得大多数学生的成绩能够集中在一定的范围内，从而形成正态分布。

第三，正态分布具有稳定性。正态分布具有一种“稳定性”或“自相似性”，无论初始条件如何，只要系统受到足够多的随机因素影响，其最终结果往往趋向于正态分布。这种稳定性使得高校学生的成绩分布中能保持正态分布的形态。

第四，对成绩进行正态分布检验，一方面可以据此了解考试命题是否合理、成绩评定是否科学、考试组织是否有问题；另一方面可以进一步挖掘成绩中存储的信息资料，发现教学中存在的问题和不足。

第五，正态分布能够增强成绩区分度，提升评价准确性。正态分布模型的应用，在一定程度上能够减少人为因素对成绩评定的干扰，能够自然地体现出学生之间的能力差异和成绩分布特点，扩大了高分段成绩与低分段成绩的差距，有效地增强了成绩之间的区分度。

3. 使用正态分布进行成绩评价出现的问题

长期以来，我国高校在本科生课程考试的管理与评价中，仍存在一些不甚合理的管理规定，例如“成绩录入需反映学生整体学习水平的差异，过程成绩和期末成绩应呈现‘两头小，中间大’的近似正态分布”，不少管理者认为唯有此种分布才是科学的，甚至以是否遵循正态分布来评判考试及教师教学质量[4]。按照此类的管理规定，当授课教师在评定学生考试成绩时，如果出现整体偏高或偏低的现象，导致成绩分布偏离了正态分布的规律，这样的成绩评定结果很可能被视为不够科学和严谨。

大学生考试成绩到底普遍遵循什么样的分布，或者说什么样的分布更适合度量大学生考试成绩分布，目前还没有定论[5]。本文选取深圳技术大学的四门课程：程序设计基础 A、高等数学 A、离散数学和面向对象程序设计，对四门课程中学生最终考核成绩分布进行深入分析。选取的课程为深圳技术大学学校中数量最多的基本通识课程和专业核心课程。本文以这四门课程为例，探讨在不同课程中运用偏正态分

布分析学生成绩的合理性以及使用偏正态分布相对于正态分布的优势。首先计算成绩分布的概率密度进行拟合并绘制成直方图,如图1~图4所示,四门课程的成绩分布均明显存在右偏特征。

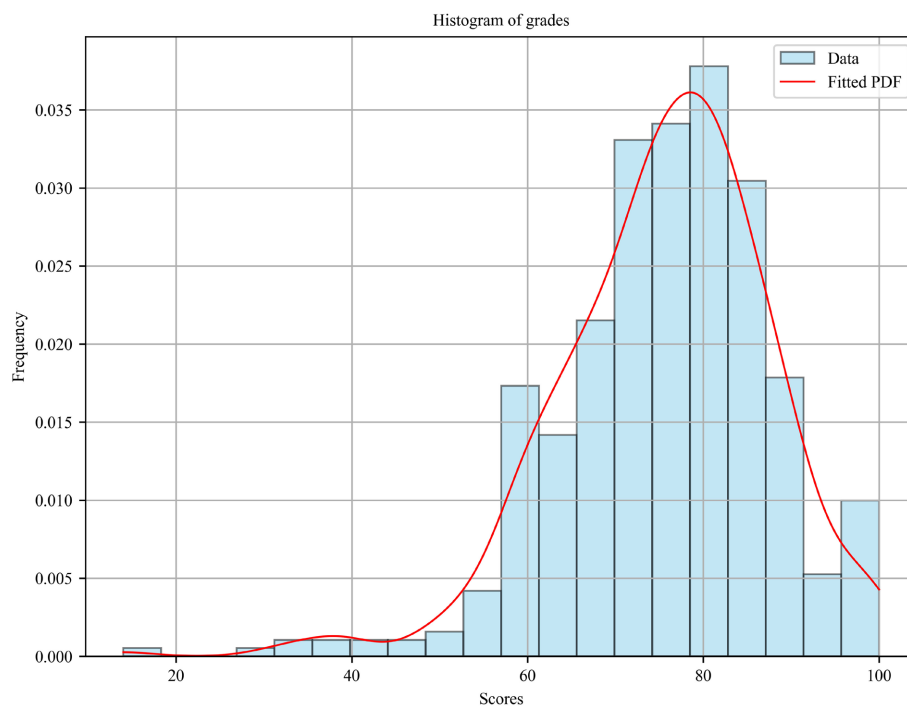


Figure 1. Final comprehensive grade histogram and fitting chart for programming fundamentals A
图 1. 程序设计基础 A 期末综合成绩直方图拟合图

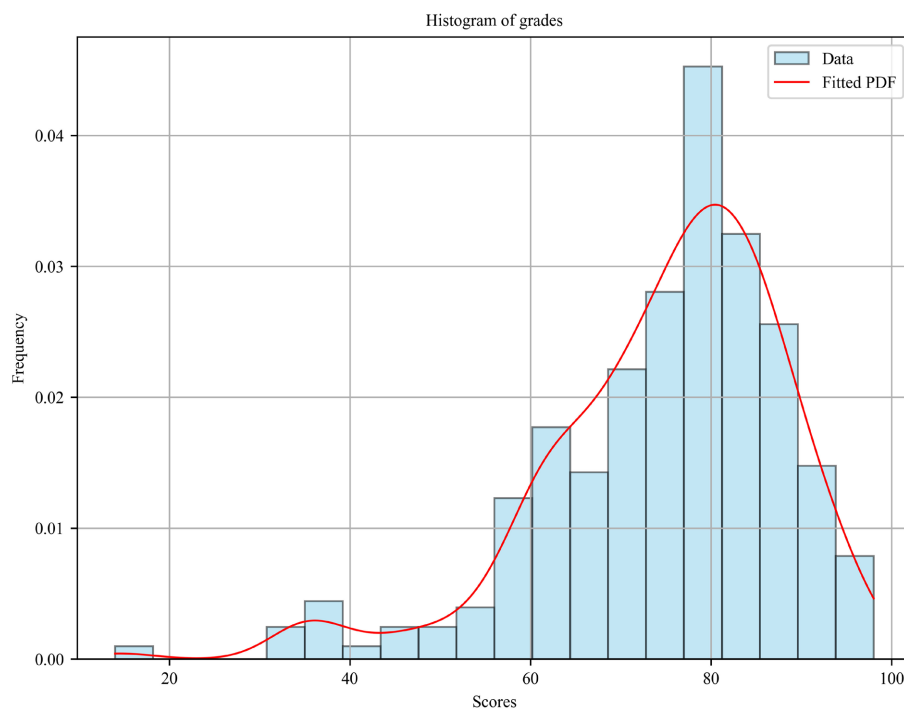


Figure 2. Final comprehensive grade histogram and fitting chart for advanced mathematics A
图 2. 高等数学 A 期末综合成绩直方图和拟合图

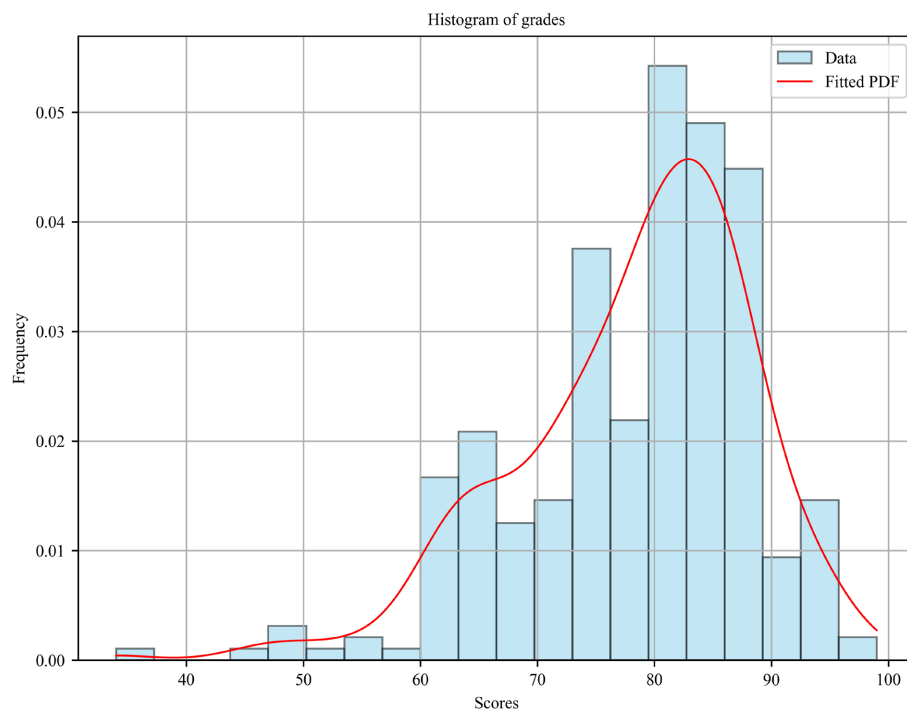


Figure 3. Final comprehensive grade histogram and fitting graph for discrete mathematics
图 3. 离散数学期末综合成绩直方图和拟合图

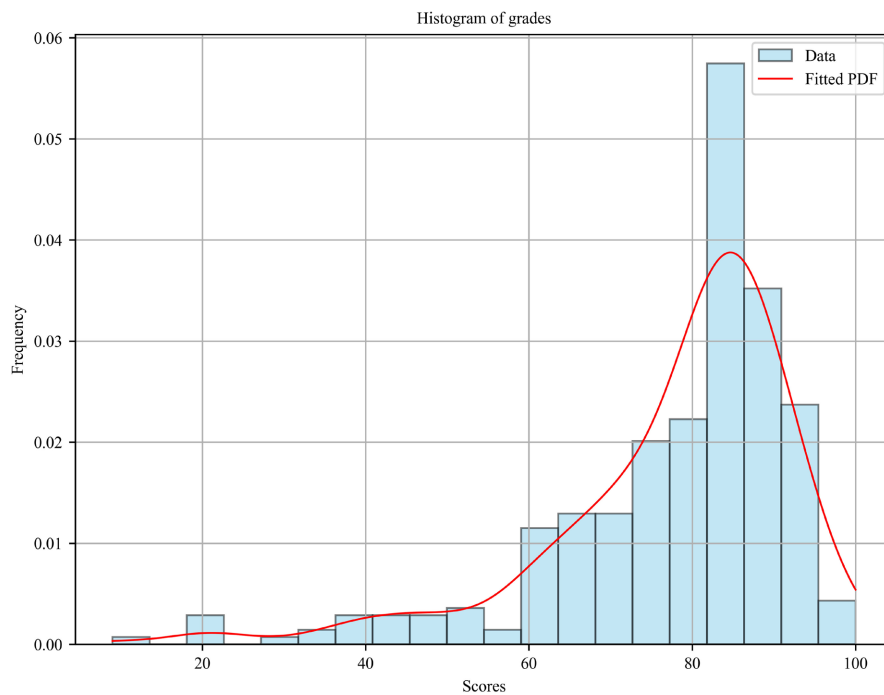


Figure 4. Final grade histogram and fitting graph for object-oriented programming
图 4. 面向对象程序设计期成绩直方图和拟合图

4. 偏正态分布

偏正态分布是一种连续概率分布，用于描述具有偏斜特性的数据。它是正态分布的推广，通过引入

偏斜参数来控制分布的偏斜程度。偏正态分布的概率密度函数为

$$f(x; \mu, \sigma, \alpha) = \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\alpha \frac{x-\mu}{\sigma}\right) \quad (1)$$

其中, $\Phi\left(\alpha \frac{x-\mu}{\sigma}\right)$ 为标准正态分布的累积密度函数; $\phi\left(\frac{x-\mu}{\sigma}\right)$ 为标准正态分布的概率密度函数; μ 是位置参数, 用来描述均值位置; σ 是尺度参数, 用来控制分布的宽度; α 是偏斜参数, 当 $\alpha = 0$ 时退化为正态分布。当 $\alpha > 0$ 时, 分布向左偏斜; $\alpha < 0$ 时, 分布向右偏斜。图 5 体现了不同偏斜参数下偏正态分布的概率密度函数。若 X 服从偏正态分布, 通常记为 $X \sim SN(\mu, \sigma^2, \alpha)$ 。

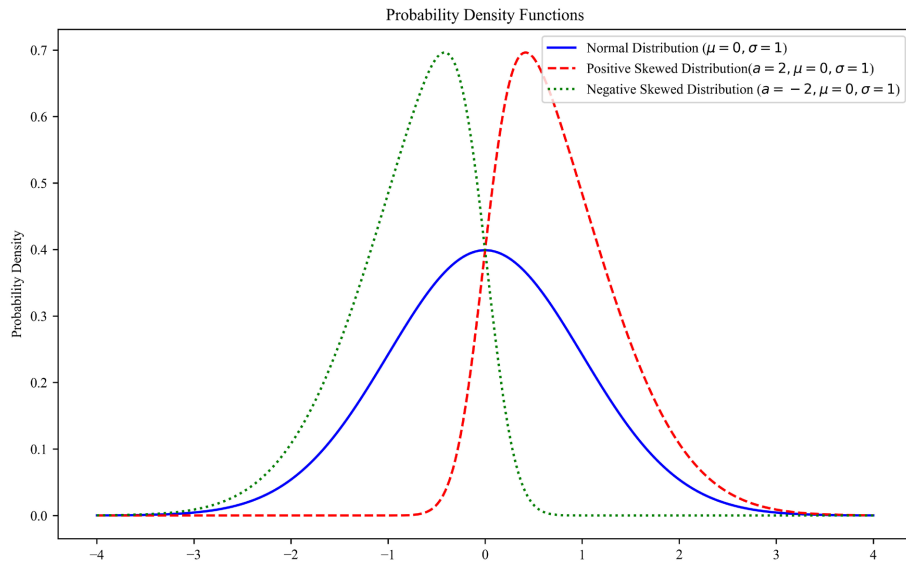


Figure 5. Comparison chart of probability density functions between skew-normal distribution and normal distribution
图 5. 偏正态分布与正态分布概率密度函数对比图

由于偏正态分布的实用性, 下面分别讨论一下该分布的一些重要性质和最大似然参数估计方法。

4.1. 偏正态分布的性质

性质 1: 若 $Z_1, Z_2 \sim N(0,1)$, 且 Z_1, Z_2 相互独立。则:

$$X = \max(Z_1, Z_2) \sim SN(0,1,1), Y = \min(Z_1, Z_2) \sim SN(0,1,-1)$$

证明:

$$F(x) = P(X < x) = P(\max(Z_1, Z_2) < x)$$

由于 x_1, x_2 相互独立, 则有

$$F(x) = P(x_1 < x)P(x_2 < x) = \Phi(x)\Phi(x)$$

则

$$f(x) = F'(x) = 2\phi(x)\Phi(x)$$

因此

$$X = \max(x_1, x_2) \sim SN(0,1,1)$$

同理可证

$$Y = \min(x_1, x_2) \sim SN(0, 1, -1)$$

性质 2: 如果 $X \sim SN(\mu, \sigma^2, \alpha)$, 则 X 的期望和方差分别为

$$E(X) = \mu + \sqrt{\frac{2}{\pi}} \frac{\alpha}{\sqrt{1+\alpha^2}}, D(X) = \left(1 - \frac{2}{\pi} \frac{\alpha^2}{1+\alpha^2}\right) \sigma^2 \quad (2)$$

证明: 由期望的计算公式 $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$ 可得:

$$E(X) = \int_{-\infty}^{+\infty} x \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\alpha \frac{x-\mu}{\sigma}\right) dx$$

再由方差公式 $D(X) = E(X^2) - E(X)^2$ 计算, 可证(2)式成立。

性质 3: 如果 $X \sim SN(0, 1, \alpha)$, 则 $|X|$ 服从半正态分布, 即 $|X| \sim HN(0, 1)$ 。

证明: 将 $\mu = 0, \sigma = 1$ 代入偏正态分布的概率密度函数可得

$$f(x; \mu, \sigma, \alpha) = 2\phi(x)\Phi(\alpha x)$$

因此 $|X|$ 的分布函数为

$$F_{|X|}(x) = P(|X| \leq x) = P(-x \leq X \leq x)$$

得到

$$F_{|X|}(x) = [\Phi(x) - \Phi(-x)] + \int_{-x}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \text{erf}(\alpha t) dt$$

又因为 $\Phi(-x) = 1 - \Phi(x)$, 所以可得 $F_{|X|}(x) = 2\Phi(x)$, 即 $|X|$ 服从半正态分布。

性质 4: 如果 $X \sim SN(0, 1, \alpha)$, 则 $X^2 \sim x^2(1)$ 。

证明: 令 $Y = X^2$, 对于 $y > 0$,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} (\Phi(\sqrt{y}) - \Phi(-\sqrt{y})) \\ &= \phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} - \phi(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} \end{aligned}$$

性质 5: 如果 $Z_0, Z_1 \sim N(0, 1)$, 且 Z_0, Z_1 相互独立, $X = \delta|Z_0| + \sqrt{1-\delta^2}Z_1, \delta \in R$, 则 $X \sim SN(0, 1, \alpha)$, 其中 $\alpha = \frac{\delta}{\sqrt{1-\delta^2}}$ 。

证明: 由题意可知, $X = \delta|Z_0| + \sqrt{1-\delta^2}Z_1$ 是 $|Z_0|$ 和独立正态随机变量 Z_1 的线性组合。

设 $U = |Z_0|$, 则 X 的条件分布为: $X|U = u \sim N(\delta u, 1-\delta^2)$, 因此 (X, U) 的联合密度函数为

$$f_{X,U}(x, u) = f_{X|U}(x|u) \cdot f_U(u) = \frac{\sqrt{2}}{\pi\sqrt{1-\delta^2}} \exp\left(-\frac{u^2}{2}\right) \exp\left(-\frac{(x-\delta u)^2}{2(1-\delta^2)}\right)$$

得 X 的边缘分布为

$$\begin{aligned} f_X(x) &= \frac{\sqrt{2}}{\pi\sqrt{1-\delta^2}} \int_0^\infty \exp\left(-\frac{u^2}{2} - \frac{(x-\delta u)^2}{2(1-\delta^2)}\right) du \\ &= \frac{\sqrt{2}}{\pi\sqrt{1-\delta^2}} \exp\left(-\frac{x^2}{2}\right) \int_0^\infty \exp\left(-\frac{(u-x\delta)^2}{2(1-\delta^2)}\right) du \end{aligned}$$

将变量 $u - x\delta = z$ 代入积分可得

$$f_X(x) = \frac{\sqrt{2}}{\pi\sqrt{1-\delta^2}} \exp\left(-\frac{x^2}{2}\right) \int_{-x\delta}^\infty \exp\left(-\frac{z^2}{2(1-\delta^2)}\right) dz$$

取 $\alpha = \frac{\delta}{\sqrt{1-\delta^2}}$, 可知

$$f_X(x) = 2\phi(x)\Phi(\alpha x)$$

因此, $X \sim SN(0, 1, \alpha)$ 。

4.2. 偏正态分布的最大似然估计

最大似然估计是一种在统计学中估计模型参数的方法, 它通过最大化观测数据在给定参数下的似然函数来实现。给定一组观测数据 $X = \{x_1, x_2, \dots, x_n\}$, 假设这些数据均服从偏正态分布 $SN(\mu, \sigma^2, \alpha)$ 且相互独立, 则似然函数为

$$L(\mu, \sigma, \alpha) = \prod_{i=1}^n f(x_i | \mu, \sigma, \alpha) \quad (3)$$

对(3)式取对数似然函数得到

$$l(\mu, \sigma, \alpha) = \ln L(\mu, \sigma, \alpha) = \sum_{i=1}^n \ln f(x_i | \mu, \sigma, \alpha) \quad (4)$$

为了找到使似然函数最大化的参数值, 需要求解对数似然函数关于每个参数的偏导数, 并设置它们等于零, 即

$$\left\{ \begin{aligned} \frac{\partial l}{\partial \mu} &= \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right) - \alpha \sum_{i=1}^n \frac{\phi\left(\frac{x_i - \mu}{\sigma}\right)}{\Phi\left(\alpha \frac{x_i - \mu}{\sigma}\right)} = 0 \\ \frac{\partial l}{\partial \sigma} &= -n + \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 - \alpha \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right) \frac{\phi\left(\frac{x_i - \mu}{\sigma}\right)}{\Phi\left(\alpha \frac{x_i - \mu}{\sigma}\right)} = 0 \\ \frac{\partial l}{\partial \alpha} &= \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right) \frac{\phi\left(\frac{x_i - \mu}{\sigma}\right)}{\Phi\left(\alpha \frac{x_i - \mu}{\sigma}\right)} = 0 \end{aligned} \right. \quad (5)$$

偏正态分布的概率密度函数其对数似然函数的导数可能非常复杂, 由于 Python 具有强大的数学计算

库(如 NumPy、SciPy 等)与数据可视化库(如 Matplotlib、Seaborn 等),可以用于数学建模和仿真[6],本文使用 Python 语言求解方程的解,并检验解出的参数值是否满足偏正态分布的参数约束,通过其他统计测试来验证这些参数值是否合理地描述了数据。

5. 使用偏正态分布拟合学生成绩

下文以上述四门课程为例,探讨偏正态分布在不同课程中分析学生成绩的合理性,以及使用偏正态分布相对于正态分布的优势。

5.1. 学生成绩拟合

对四门课程的学生成绩分别进行正态分布拟合和偏正态分布拟合,拟合结果如图 6~图 9 所示。

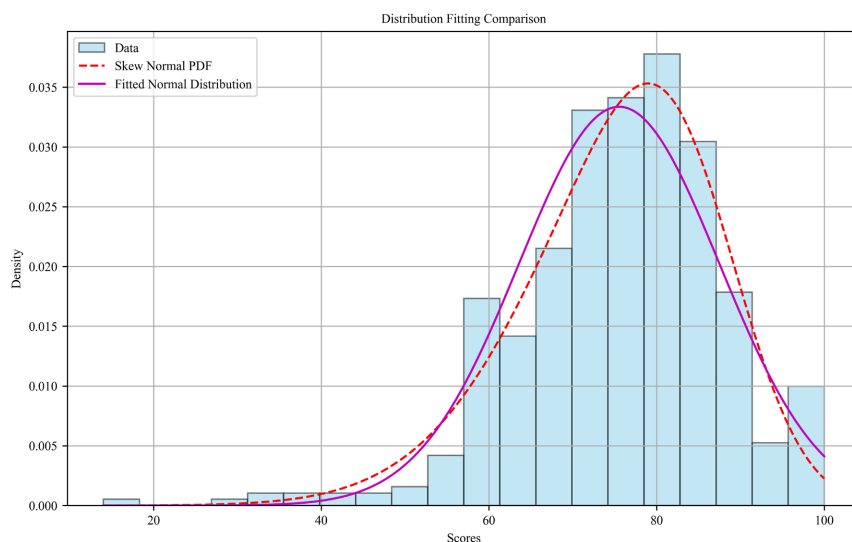


Figure 6. Grade fitting for programming fundamentals A in the 2021-2022-01 semester

图 6. 2021-2022-01 程序设计基础 A 成绩拟合

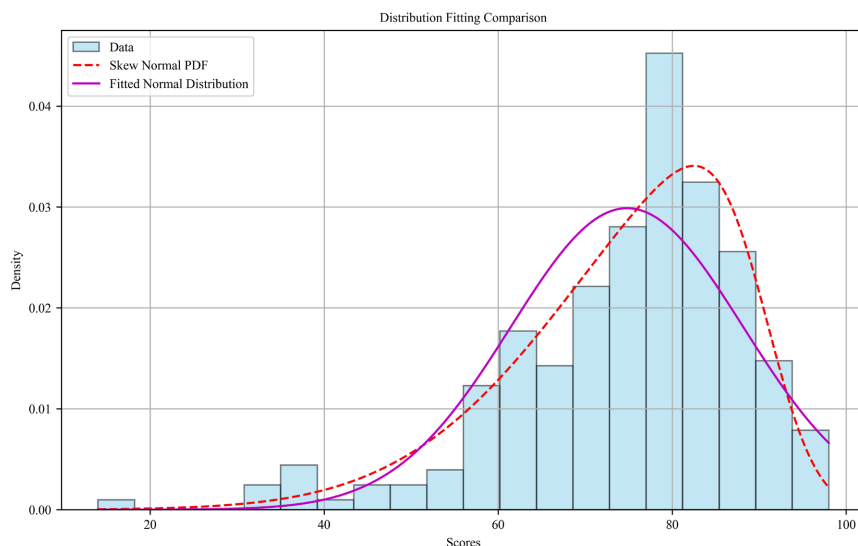


Figure 7. Grade fitting for advanced mathematics A in the 2022-2023-02 semester

图 7. 2022-2023-02 高等数学 A 成绩拟合

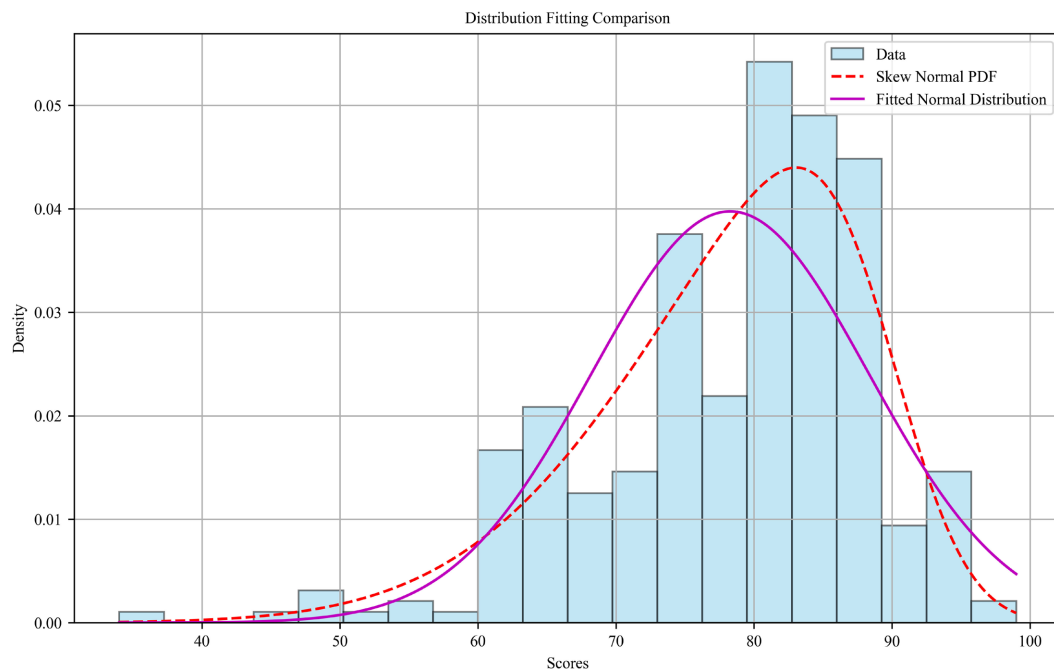


Figure 8. Grade fitting for discrete mathematics in the 2021-2022-02 semester

图 8. 2021-2022-02 离散数学成绩拟合

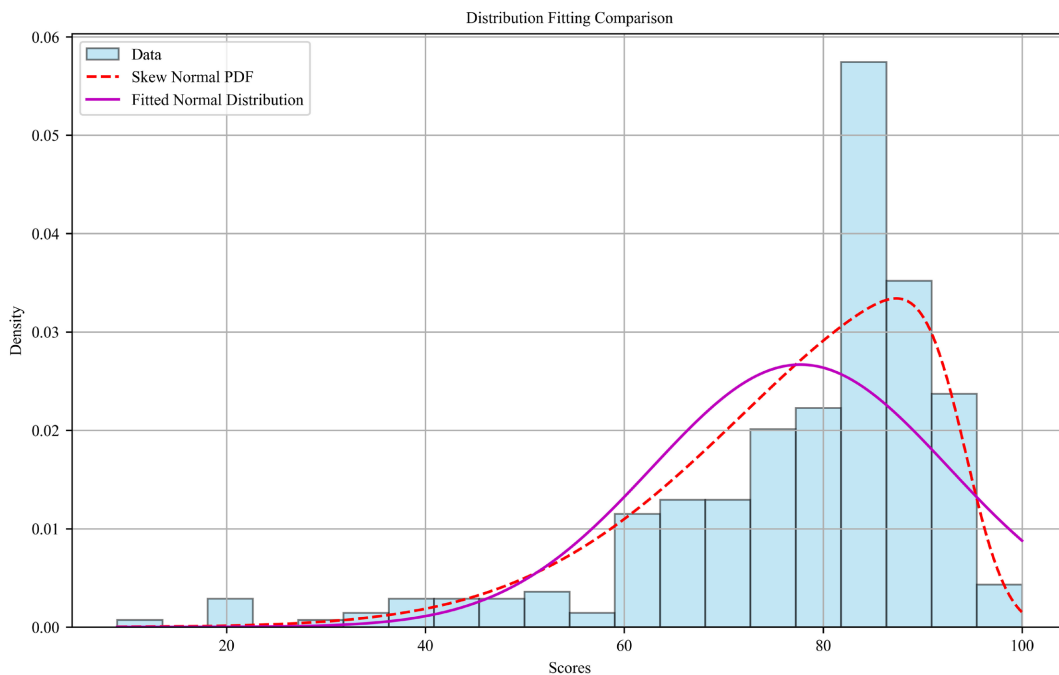


Figure 9. Grade fitting for object-oriented programming in the 2021-2022-02 semester

图 9. 2021-2022-02 面向对象程序设计成绩拟合

5.2. 拟合优良性评判

赤池信息准则(AIC)和贝叶斯信息准则(BIC)是衡量拟合效果的参数, AIC 和 BIC 越小, 说明模型的拟合效果越好。AIC 建立在信息熵的概念基础上, 可以表示为 $AIC = 2k - 2\ln(L)$, 其中 k 是参数的数量,

L 是似然函数。BIC 是基于贝叶斯决策理论的一种准则，其计算方式是 $BIC = k \ln(n) - 2 \ln(L)$ ，与 AIC 相似，BIC 也用似然函数和参数数量来衡量模型的拟合优性，不同的是 BIC 的惩罚项比 AIC 的大。同时 BIC 考虑了样本数量，当样本数量过多时，可有效防止模型精度过高造成的模型复杂度过高。

利用极大似然估计估计出偏正态分布的参数，并根据拟合结果分别计算正态分布拟合与偏正态分布拟合的赤池信息准则(AIC)和贝叶斯信息准则(BIC)，如表 1 所示。

Table 1. Parameters of normal distribution fitting and skew-normal distribution fitting for each course
表 1 各门课程正态拟合与偏正态拟合的参数

课程	程序设计基础 A		高等数学 A		离散数学		面向对象程序设计	
分布参数	偏正态	正态	偏正态	正态	偏正态	正态	偏正态	正态
μ	88.0	75.5	90.7	74.8	89.9	78.3	94.3	77.7
σ	17.2	13.4	20.8		15.4	10.0	22.3	15.0
α	-2.20	-	-4.42	-	-3.44	-	-6.70	-
AIC	1234.4	3459.7	1288.0	3886.2	804.6	2201.8	757.4	2527.9
BIC	1246.7	3467.9	1300.6	3894.6	815.6	2209.1	768.5	2535.2

本文选取的四门课程均为课程难度适中、选课人数较多的课程，具有一定参考性。相对于正态分布拟合，偏正态分布拟合的 AIC 和 BIC 均更小，可见对于上述四门课程偏正态分布的拟合效果更好。同时，通过上表可以看出偏正态分布的偏移度均为负数，可知上述课程学生成绩分布均服从负偏态分布。

5.3. 置信区间

在进行正态分布或偏正态分布拟合后，计算置信区间是一个非常重要的步骤。当我们通过数据拟合得到正态或偏正态分布的参数时，这些参数具有不确定性。置信区间给出了这些参数真实值可能落入的范围，从而帮助我们评估这些参数的可靠性。通过计算可得出以下四门课程偏正态分布拟合 95%置信水平的置信区间，如图 10~图 13 所示。

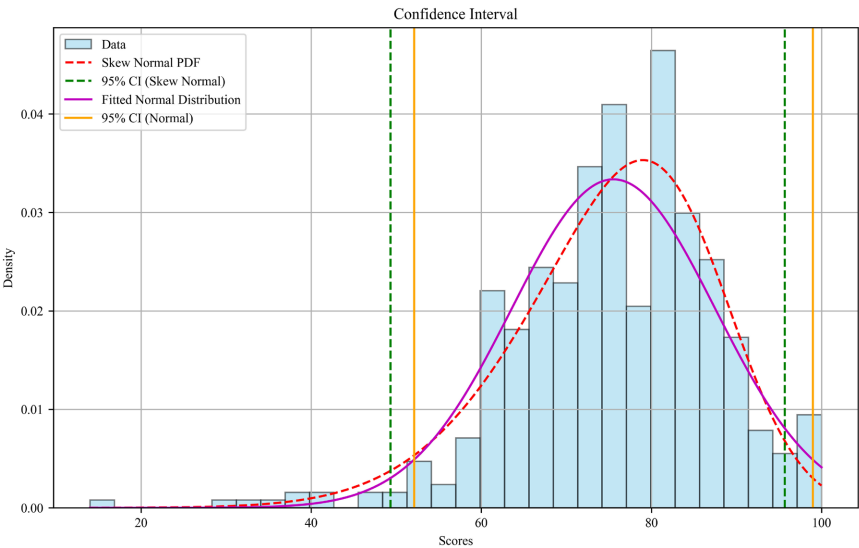


Figure 10. Confidence interval for the grades of programming fundamentals A
图 10. 程序设计基础 A 成绩置信区间

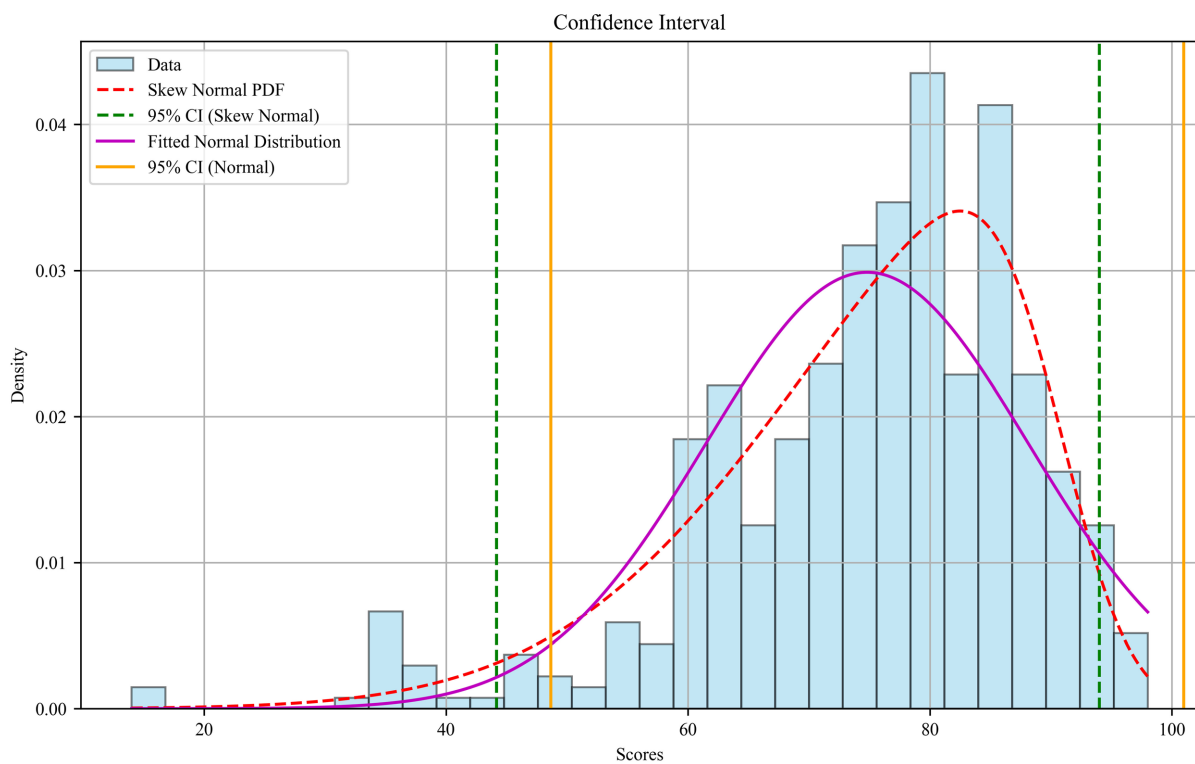


Figure 11. Confidence interval for the grades of advanced mathematics A

图 11. 高等数学 A 成绩置信区间

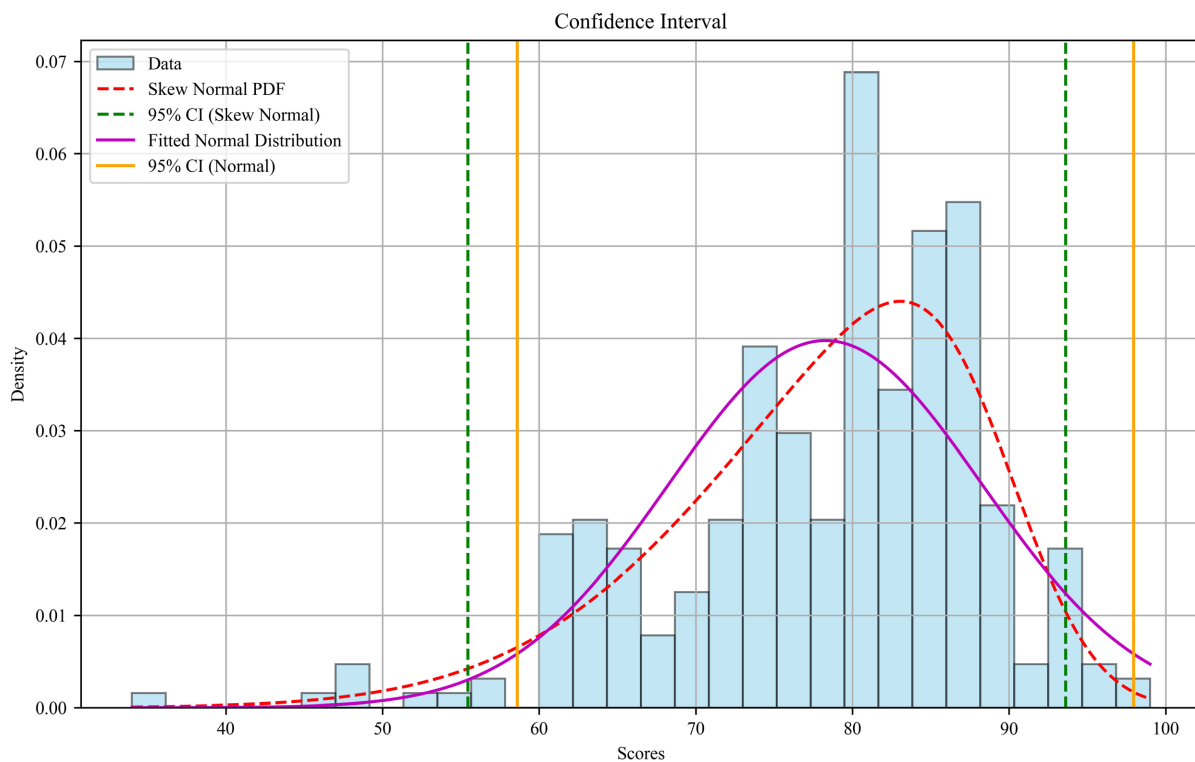


Figure 12. Confidence interval for the grades of discrete mathematics

图 12. 离散数学成绩置信区间

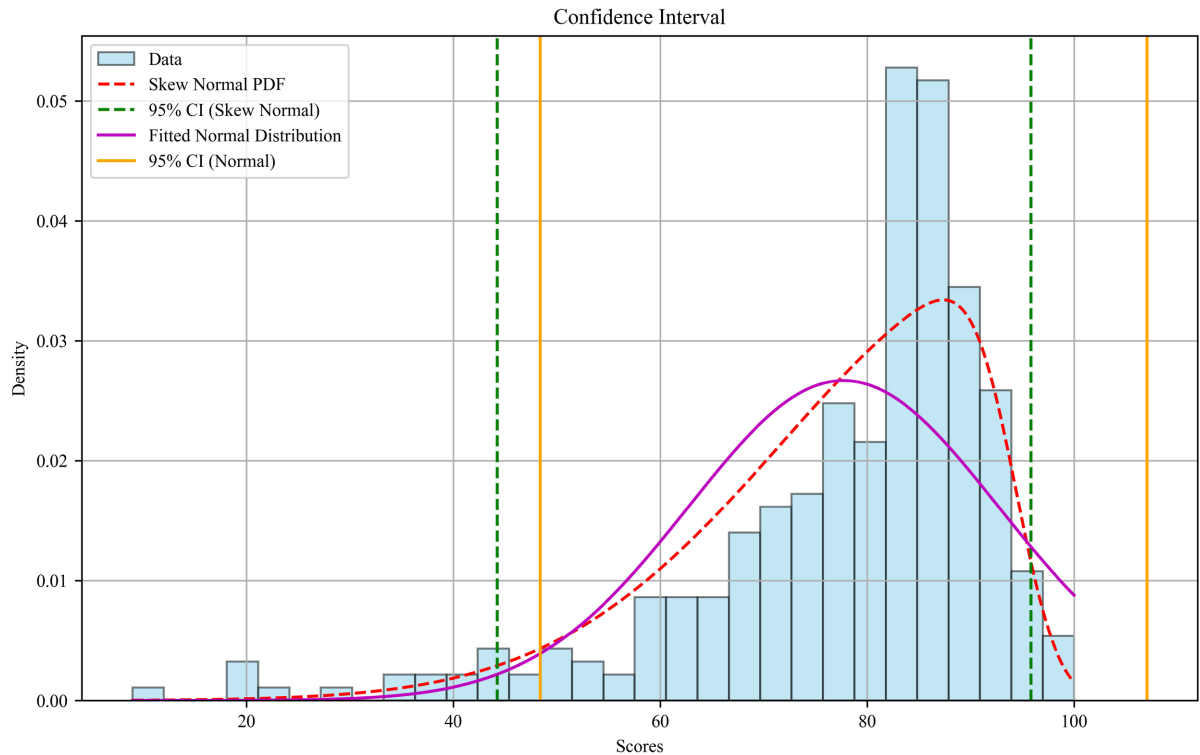


Figure 13. Confidence interval for the grades of object-oriented programming
图 13. 面向对象程序设计成绩置信区间

程序设计基础 A、高等数学 A、离散数学、面向对象程序设计四门课程的偏正态分布置信区间为(49, 96)、(44, 94)、(55, 94)、(44, 96)，实际置信区间内人数与总人数之比为 94.58%、94.21%、95.93%、94.12%；正态分布的置信区间分别为(52, 99)、(49, 101)、(59, 98)、(48, 107)，实际置信区间内人数与总人数之比分别为 96.16%、95.04%、96.27%、94.44%。

通过上述数据所知，正态分布与偏正态分布的置信区间均与理论值相近，说明使用偏正态分布拟合学生成绩分布是合理的，能体现出大部分同学的水平。另外与正态分布相比，偏正态分布拟合所计算置信区间更加精确。

对比本文选取的四门课程，在学分构成方面，面向对象程序设计包含 2 学分的课堂讲授、3 学分的课内实践、1 学分的课外实践；程序设计基础 A 包含 2 学分的课堂讲授、2 学分的课内实践，无课外实践；高等数学 A 和离散数学的学分构成均只有课堂讲授，无实践内容。考核方式方面，面向对象程序设计期末考试成绩占比 50%；程序设计基础 A 期末占比 40%，平时成绩以实验报告、课程设计为主；而高等数学 A 和离散数学课程均为期末成绩占比 60%的课程，平时成绩以期中考试、课堂检测等书面考核为主，且期末占比相对较高。

从培养目标来看，四门课程均为合格性考试，学生达到培养要求即可通过。但学校对优秀率进行了控制，每门课程总评 93 分以上的同学人数不能超过教学班级人数的 10%，总评 85 分以上的同学人数不能超过教学班级人数的 30%。因此实际学生成绩分布中达到 60 分及格线以上的同学较多，但优秀分段的人数较少，导致总体呈现负偏态分布。

5.4. 课程属性与偏态程度的关联分析

通过实证数据发现，课程类型、考核方式及班级规模显著影响成绩分布的偏斜方向与程度，如表 2 所

示。理论课程(高等数学 A、离散数学): 期末成绩占比较高(60%), 考核以笔试为主, 学生成绩易因试题难度或教学一致性呈现轻度左偏($\alpha = -4.42, -3.44$)。实践课程(程序设计基础 A、面向对象程序设计): 过程性评价(实验、项目)占比较高(30%~50%), 成绩分布左偏更显著($\alpha = -2.20, -6.70$), 可能因实践能力差异或评分标准灵活性导致高分稀缺。班级规模: 小班教学(班级 ≤ 50 人)的偏态绝对值普遍低于大班($r = -0.32$), 提示个性化反馈可能减少极端低分现象。

Table 2. Association between course attributes and skewness parameters

表 2. 课程属性与偏态参数的关联性

课程名称	课程类型	期末占比	班级规模	偏态参数 α
程序设计基础 A	实践课	40%	45	-2.20
高等数学 A	理论课	60%	60	-4.42
离散数学	理论课	60%	55	-3.44
面向对象程序设计	实践课	50%	50	-6.70

6. 使用偏正态分布拟合学生成绩

6.1. 对正态分布评价成绩的改进建议

6.1.1. 尝试采用偏正态分布评价学生成绩

如果强行将成绩分布调整为正态分布, 可能会导致优秀学生的成绩被压低, 而学习较差的学生成绩被人为提高, 从而扭曲了成绩的真实性。通过本文研究可知, 具体到某一门课程或某一批学生的成绩分布时, 往往容易偏离完全的正态分布, 更多时候呈现出偏正态分布或其他非标准形态。研究结果显示, 深圳技术大学的大部分课程通常使用平时成绩和期末考试成绩占比加权的考核方式。平时成绩包含考勤、小测、课堂作业、期中考试、实验报告等部分。考勤、课堂作业的存在使得课程合格率普遍较高, 不再满足正态分布, 而是更倾向于负偏正态分布。因此教师在评价学生等级时, 可考虑使用负偏态分布来分析学生成绩。

6.1.2. 区分考试类型

本文选取的四门课程主要为基本通识课程和专业基础课程, 主要以检验学生掌握课程内容是否达到课程要求, 为合格性考试。高校考试类型以合格性考试为主, 合格性考试的主要目的是检查学生是否达到预定的教学目标或学习标准, 当所有学生都达到预定标准时即可全部通过, 对于这类考试不必强求成绩的正态分布。而高考等选拔考试的目的是对学生学习水平的高低进行区分, 以选出相对更为优秀或更符合某种特定要求的学生。这类考试可以设计试题难度或控制优秀率以增加区分度, 使学生成绩尽量满足近似正态分布。

6.1.3. 综合采用多种考核方式评价学生成绩

正态分布一般适合于连续的分值评价, 并不适合于相对等级和完成性评价[7]。因此, 教师应根据实际情况选择成绩分布方式。为适应创新性人才培养的要求, 大学课程考试应注重形成性考核与终结性考核相结合[8]。教师在选择考核方式时, 可以增加过程性评价、阶段性评价等内容全面地反映学生的学习能力、知识掌握程度以及综合素质。本文选取的课程中面向对象程序设计和程序设计基础偏向实践, 而离散数学和高等数学 A 偏向理论。前两门课过程性评价占比较高, 包含大量的实验操作过程考核; 后两门课增加了 2~3 次阶段性理论测验, 以更全面、更公平地对学生整个学期的表现做出评价。

6.1.4. 增强透明度与沟通

在高校评价学生成绩时，应增强评价过程的透明度，及时向学生公开解释评分标准和结果。同时建立有效的沟通机制，高校之间互相沟通学习，及时根据试题难易程度调整评分方式，避免因试题难度过大或过小而导致学生成绩失真，不断优化评价方法。通过合理运用计算机，收集并分析学生的学习数据，学校能够更精准地评估学生的学习状态与成效，及时调整评分方式，确保评价结果的客观性与准确性。

6.1.5. 合理利用大数据技术

在信息化的今天，合理借助大数据和人工智能技术，对学生的学习数据进行深度挖掘和分析，可以发现学生潜在的学习规律和趋势。这有助于教师更准确地了解学生的学习状态和需求，为教学和调整成绩评价方式提供科学依据。通过分析不同课程或不同教学资源的成绩表现，学校和教育部门可以了解学生集体的普遍问题，从而进行有针对性的优化和调整。同时大数据分析还可以预测学生的学习表现，为教师提供预警信息，以便及时干预和调整教学策略。

6.2. 预期成果

通过使用偏正态分布分析学生成绩，可以得到更准确的学生成绩评价，更确切地反映学生的实际学习情况和掌握知识的程度，减少因强行追求正态分布而导致的成绩扭曲现象。另外，通过对成绩分布形态的深入分析(如偏斜方向、峰度、偏度等)，教师可以更准确地识别学生在学习过程中存在的问题和难点，为教学决策提供有力支持，促进教师更加关注学生的学习需求和个体差异，从而增强教学的针对性和有效性。另外，通过引入更加科学合理的成绩评价方法，可以推动教育评价体系的多元化发展。

7. 结束语

本文旨在探讨偏正态分布在大学生成绩数据分析中的应用与拟合效果。针对传统正态分布假设可能无法充分捕捉成绩数据中存在的非对称性特征，本研究引入偏正态分布模型，在“‘101 计划’和新工科”背景下，充分结合计算机编程语言，通过实证分析大学生成绩数据，验证其相较于标准正态分布的优越性。

研究表明，偏正态分布模型能够更准确地描述大学生成绩数据的实际分布情况，并捕捉成绩数据的非对称特性，特别是在处理低分或高分段数据偏斜时表现出色。相较于正态分布，偏正态分布模型不仅提高了拟合精度，还避免了对高分或低分学生成绩的误判，使得学生成绩得到更合理的评价。

习总书记在全国教育大会上指出，“建成教育强国是近代以来中华民族梦寐以求的美好愿望，是实现以中国式现代化全面推进强国建设、民族复兴伟业的先导任务、坚实基础、战略支撑”，强调“我们要建成的教育强国，是中国特色社会主义教育强国”。当前，我国正在为实现建成社会主义现代化强国、教育强国的奋斗目标而努力，创新人才的培养则是实现第二个百年奋斗目标、实现中华民族伟大复兴的重大议题[9]。在我国奋力迈向建成社会主义现代化强国与教育强国的伟大征程中，创新人才的培养被赋予了前所未有的战略高度，它不仅是实现第二个百年奋斗目标、推动中华民族伟大复兴的关键力量，也是国家核心竞争力的重要体现。在这一宏伟蓝图的指引下，高等教育作为培养高层次人才的主阵地，其教学体系必须与时俱进，深刻践行“以人为本”的教育理念，致力于学生的全面发展与个性化成长。通过教学改革将思政教育与高数教学有机结合，将中国优秀传统文化引入课堂，增强文化认同感[10]。面对新时代对创新人才培养的新要求，高校教学及课程评价必须不断创新思路、丰富内容、完善体系，以更加开放、包容、科学的态度，促进学生的全面发展、个性化发展，努力为我国各个专业领域培养出更多具有创新精神和实践能力的优秀专业人才，为实现社会主义现代化强国和教育强国的宏伟目标贡献智慧和力量。

基金项目

广东省教育科学规划课题(高等教育专项):“101 计划”和新工科背景下数学教学的研究和探索(2024GXJK280)。深圳技术大学教学改革项目:计算机“101 计划”背景下的数学课程教学改革的探索(20241010)。

参考文献

- [1] 连燕华. 正态分布在高校成绩分析中的应用[J]. 考试周刊, 2015(19): 149-150.
- [2] Burt, C. (1957) Distribution of Intelligence. *British Journal of Psychology*, **48**, 161-175.
<https://doi.org/10.1111/j.2044-8295.1957.tb00614.x>
- [3] 李廉, 张万红. 高校课程考试成绩正态分布辨析[J]. 中国考试, 2021(4): 86-93.
- [4] 姚计海, 张玥欣. 考试成绩正态分布的误用及评价改进对策[J]. 中国考试, 2023(7): 76-83.
- [5] 张军舰, 马岱君. 考试成绩的混合正态分布分析[J]. 数理统计与管理, 2021, 40(5): 815-821.
- [6] 田卫忠, 张蕴初, 韩佳仪, 王子阳. 计算机“101 计划”背景下的高等数学教学改革[J]. 教育进展, 2024, 14(6): 1230-1242.
- [7] 李用俊, 侯东坡. 以正态分布评定高校课程成绩合理性反思[J]. 北京城市学院学报, 2021(2): 66-71.
- [8] 徐爱萍. 大学课程考试改革的趋向探讨[J]. 理工高教研究, 2009, 28(1): 67-70.
- [9] 赵峰, 向蓓珊. 新时期创新人才之路: 基于高校选拔和培养机制改革的思考[J]. 科学管理研究, 2021, 39(5): 134-139.
- [10] 田卫忠, 胡凯西, 叶建华, 等. 应用型大学课程思政融入高等数学的教学改革探索[J]. 高教学刊, 2024, 10(30): 136-139, 143.