Vol 17, Issue 63, 2026, 1-39 ISSN: 2252-004x
eISSN: 2470-6240
Article Type: Research Article
Received in revised form July 20, 2025
Accepted January 12, 2026
Received April 18, 2025
Published Online March 26, 2026

Educational Measurement
https://jem.atu.ac.ir/

ATU PRESS

A Comparison of the Current National University Entrance Exam Ranking Method in Iran (Linear Combination of Standardized Scores) with Principal Component Analysis and the Mahalanobis Distance

Mahsima Arbabi — Ph.D. Student in Assessment & Measurement, University of Tehran, Tehran, Iran. E-mail: mahsima.arbabi123@gmail.com

Ebrahim Khodaie * — *Corresponding Author*, Professor, Department of Curriculum Development and Instruction Methods, University of Tehran, Tehran, Iran. E-mail: khodaie@ut.ac.ir

Keyvan Salehi — Associate Professor, Department of Educational Psychology, University of Tehran, Tehran, Iran. E-mail: keyvansalehi@ut.ac.ir

Abstract

Ranking candidates in Iran's National University Entrance Exam (Konkoor) is a critical step in the university admission process. Traditional ranking methods, based on a linear combination of standardized scores, face limitations such as neglecting data distribution and inter-variable correlations. This study aims to enhance the ranking process by employing Principal Component Analysis (PCA) for dimensionality reduction and Mahalanobis Distance for candidate ranking. The study population consisted of 126,728 candidates from the Mathematics and Technical Sciences group in the 2021 exam, reduced to 110,550 after eliminating missing data. Data were preprocessed, subjected to PCA for dimensionality reduction, and ranked using Mahalanobis Distance. Analyses were conducted using SPSS and Python. The results indicate that the proposed method offers greater discriminatory precision in distinguishing among candidates while retaining critical informational variance. This enhanced approach facilitates a more effective identification of academically superior candidates for university admission. The findings suggest that integrating Principal Component Analysis and the Mahalanobis Distance provides a robust and more effective methodological framework for candidate ranking, with the potential to significantly improve upon current admission methodologies.

Keywords: National University Entrance Exam, linear combination of standardized scores, principal component analysis, multivariate ranking, Mahalanobis distance

Cite this Article: Arbabi, M., Khodaie, E., & Salehi, K. (2026). A Comparison of the Current National University Entrance Exam Ranking Method in Iran (Linear Combination of Standardized Scores) with Principal Component Analysis and the Mahalanobis Distance. *Educational Measurement*, *17*(63), 1-39. https://doi.org/10.22054/jem.2025.85317.3608

© 2016 by Allameh Tabataba'i University Press
Publisher: Allameh Tabataba'i University Press

Homepage: https://jem.atu.ac.ir/

Introduction

Ranking individuals within multivariate datasets remains a persistent challenge in educational assessment and related fields. Traditional methods, such as linear ranking and weighted averages, have been widely used to compare individual performance (Njoroge et al., 2014). However, these approaches often fail to account for the influence of correlated variables and the underlying data distribution, which can lead to significant inaccuracies in ranking (Dorans, 2004). In Iran's National University Entrance Exam (Konkoor), candidate ranking is based on a linear combination of standardized scores from general and specialized subjects, further adjusted by academic records. Raw scores for each subject are first calculated based on the number of correct and incorrect responses. These raw scores are then standardized to facilitate comparison across all candidates. The final composite score is derived by combining these standardized scores using subject-specific weights predetermined for each academic subgroup. Academic records contribute an additional 5% to 30% to this composite score, depending on the candidate's diploma type and chosen exam group (National Organization for Educational Testing, 2021).

The standardized score, while conceptually analogous to a Euclidean distance metric, fails to account for inter-subject correlations. This omission renders it sensitive to outliers and the specific distribution of the data. In contrast, advanced statistical methods such as Principal Component Analysis (PCA) and Mahalanobis Distance incorporate the covariance structure between variables, thereby providing a more robust and accurate ranking framework (Holland & Wainer, 2012). PCA reduces data dimensionality while preserving essential information and maximizing variance (Jolliffe & Cadima, 2016), whereas Mahalanobis Distance measures multivariate divergence from a centroid by accounting for the inherent covariance structure (De Maesschalck et al., 2000). These techniques have been widely applied in educational research to evaluate student performance multidimensionally and enhance the objectivity of decision-making processes.

Research Questions

This study addresses the following research questions:

(1) Does the integration of Principal Component Analysis (PCA) and Mahalanobis Distance improve the precision and differentiation of

candidate ranking compared to the traditional method of a linear combination of standardized scores?
(2) How effectively does this combined methodological approach mitigate the impact of outliers and better account for inter-subject performance correlations?
The study aims to conduct a comparative analysis using real examination data to address these questions.

Literature Review

Principal Component Analysis (PCA) has been extensively applied in educational research for dimensionality reduction and for identifying latent factors underlying student performance. Njoroge et al. (2014) demonstrated that PCA can highlight critical variables, such as English language proficiency, that are predictive of overall academic success. Similarly, Twenefour et al. (2015) found PCA to provide a more precise analysis of student performance than traditional grading systems. In a different application, Nunes et al. (2015) utilized PCA to assess academic management efficacy within engineering programs, while Aboagye and Mensah (2016) employed it to differentiate key performance criteria in mathematics and statistics education. More recently, Tsoni et al. (2021) integrated PCA with clustering techniques to analyze behavioral patterns in distance learning, and Raj and Vidyaa Thulasiraman (2021) used PCA as a preprocessing step to improve the efficiency and personalization of e-learning systems.

Mahalanobis Distance is a robust multivariate metric that addresses key limitations of traditional ranking methods by incorporating the covariance structure of the data. De Maesschalck et al. (2000) demonstrated its efficacy in detecting subtle group differences. Within educational research, Giovannella et al. (2013) employed it to evaluate the impact of Web 2.0 tools on digital learning ecosystems. Singh (2022) and Kumar and Thakur (2019) applied the metric to rank universities based on multidimensional performance indicators such as graduation rates and research output. Arai et al. (2021) enhanced the prediction of language proficiency by integrating Mahalanobis Distance with regression modeling. More recently, Ahmed et al. (2024) utilized it to analyze shifts in student performance patterns during the COVID-19 pandemic.

The integration of Principal Component Analysis (PCA) and Mahalanobis Distance offers a novel and robust approach to

Homepage: https://jem.atu.ac.ir/

multivariate ranking. Twenefour et al. (2015) noted that PCA simplifies data complexity and reduces multicollinearity, thereby enabling Mahalanobis Distance to assess statistical distances more effectively within a transformed, orthogonal space. Mehra and Agrawal (2021) reported that this combination significantly enhances predictive accuracy and improves the early identification of at-risk students. However, the application of this integrated methodology to candidate ranking in Iran's Konkoor exam remains underexplored, a gap that underscores the necessity and novelty of the present study.

1. Maximum Likelihood Estimation in Adaptive Testing: This method estimates student ability independently of raw score scales, offering high reliability in computerized adaptive testing environments (Padmaji et al., 2015).
2. Peer-Based Ranking: This approach involves students evaluating their peers' work, which fosters critical thinking and reduces the instructor's grading burden. Algorithms such as Hyperlink-Induced Topic Search (HITS) ensure fairness by weighting evaluations based on the credibility of the evaluator (Overveld & Verhoeff, 2013).
3. Stochastic Gradient Descent for Large-Scale Educational Data: This technique efficiently manages sparse, high-dimensional educational data. It is often employed with Tikhonov regularization to prevent model overfitting and to optimize pairwise ranking loss functions (Taş, 2022).
4. Human-Centered AI for Interactive Learner Profiles: AI-driven analysis of student response data is used to generate dynamic, interactive learner profiles. These profiles facilitate targeted educational interventions and personalized learning pathways (Guo et al., 2024).
5. Extended Multi-Attribute Ranking: This methodology evaluates students holistically based on both theoretical knowledge and practical skills. It employs multi-attribute utility functions to generate a balanced and comprehensive assessment (Borissova & Keremedchiev, 2019).
6. Ranking in Iran's National Entrance Exam (Konkoor): Raw scores are calculated from the number of correct and incorrect responses. These scores are then standardized to ensure comparability across all examinees. The final ranking score is a composite derived from combining these standardized scores with a percentage drawn from the student's academic records. Candidates are ranked within their specific

Homepage: https://jem.atu.ac.ir/

academic subgroup, with subject weights predetermined by discipline (National Organization for Educational Testing, 2021).

7. Mahalanobis Distance in Ranking Systems: Unlike Euclidean distance, Mahalanobis Distance incorporates the covariance structure of the data, providing a robust statistical metric for analyzing correlated variables within complex datasets. Its applications include ranking MBA candidates based on multidimensional profiles (Ketkar & Vaidya, 2014) and detecting anomalies in academic performance patterns (Lim & Lanckriet, 2014).

Advantages of Mahalanobis Distance: This metric provides a robust statistical framework that accounts for inter-variable correlations. It demonstrates resilience to outliers, particularly when enhanced with shrinkage estimators, and maintains computational efficiency even with large-scale datasets (Cheng et al., 2022; Behrendt et al., 2024).

Challenges: Its efficacy is contingent upon a precise analysis of data variability and covariance structures. A key limitation is its underlying assumption of multivariate normality, a condition that is often violated in real-world educational data (Lim & Lanckriet, 2014).

Methodology

This descriptive-analytical study utilized data from 126,728 candidates in the Mathematics and Technical Sciences discipline of the 2021 National University Entrance Exam. After the removal of cases with missing data, the final analytical sample consisted of 110,550 candidates. The dataset comprised standardized scores from seven subjects—Persian Language, Arabic Language, Islamic Studies, Foreign Language, Mathematics, Physics, and Chemistry—as well as each candidate's high school diploma academic records. All data were sourced from the National Organization for Educational Testing.

Analysis Method

Principal Component Analysis (PCA) with Promax rotation was employed to reduce data dimensionality, transforming the original standardized scores and academic records into two principal components that preserved the maximum variance. Promax rotation was selected to allow for potential correlations among components, thereby enhancing the interpretability of the resulting factor structure (Adachi, 2004). The component scores were summed for each candidate, with the highest aggregate score identifying the top-

Homepage: https://jem.atu.ac.ir/

performing individual. The Mahalanobis Distance from this top performer was then calculated within the two-dimensional component space, incorporating the covariance matrix to account for the underlying data structure. Candidates were ranked based on these distances, with shorter distances corresponding to higher ranks. All analyses were performed using SPSS (version 21) and Python (version 3.9.7). The proposed ranking method was subsequently compared against the traditional linear combination approach using several metrics, including rank correlation coefficients, the magnitude of rank changes, and the frequency of tied ranks.

Traditional Ranking in Iran's Exam

Raw scores are calculated as:

$$\textbf{Raw Score} = \frac{\textbf{3Number of Correct Answers} - \textbf{Number of Incorrect Answers}}{\textbf{3Total Questions}} \times \textbf{100}$$

Raw scores are converted to normalized scores through a three-step process:

1. The relative frequency distribution of the raw scores is calculated.
2. The percentile rank of each raw score is determined based on its relative frequency and the cumulative distribution of scores.
3. The standardized (Z) score is derived by applying the inverse cumulative distribution function (quantile function) of the standard normal distribution to the percentile rank.

$$\Phi(z) = \frac{\hat{Q}(y)}{100} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-\omega^2}{2}} d\omega$$

Where:

- Φ(z): The cumulative distribution function of the standard normal distribution.
- ω: The integration variable, ranging from $-\infty$ to z.
- $\hat{Q}(y)$: The percentile rank of the score.

Finally, the normalized scores are converted to standardized test scores with a mean of 5000 and a standard deviation of 2500 using the following formula:

sc=σ(sc)z+μ(sc)

Where:

- scscsc: The standardized score.
- σ(sc): The standard deviation of the standardized scores (2500).

Homepage: https://jem.atu.ac.ir/

- μ(sc): The mean of the standardized scores (5000).
- z: The standard normal score corresponding to the percentile rank.

The candidate's total score is calculated as a weighted linear combination of standardized scores from both general and specialized subjects. The weights are determined by the officially prescribed coefficients for the candidate's respective academic subgroup (National Organization for Educational Testing, 2021).

Proposed Method

PCA with Promax rotation extracted two components, preserving 83.56% of variance. Mahalanobis Distance was calculated as:

$$\boldsymbol{D_M} = \sqrt{(\boldsymbol{x} - \mu)^{\boldsymbol{T}} \boldsymbol{S^{-1}} (\boldsymbol{x} - \mu)}$$

where x is the observation vector, μ is the mean vector, and S is the covariance matrix. This approach reduced computational complexity and improved ranking precision by focusing on maximum performance rather than mean-based metrics (Yin, 2023).

Results

Data analysis was conducted on a sample of 110,550 candidates from the 2021 Mathematics and Technical Sciences discipline. The Kaiser-Meyer-Olkin (KMO) measure yielded a value of 0.916, indicating superb sampling adequacy for factor analysis. Bartlett's test of sphericity was significant ($\chi^2(28) = 323{,}561.658$, $p < .001$), confirming that the correlations between variables were sufficiently strong for the analysis. Principal Component Analysis (PCA) extracted two components, which collectively explained 83.56% of the total variance (Component 1: 62.84%; Component 2: 20.72%). Following Promax rotation, these components were interpreted and labeled as "General Skills and Basic Sciences" and "Mathematical and Computational Abilities."

The candidate with the highest sum of component scores was selected as the multivariate reference point. Mahalanobis distances from this ideal candidate were calculated for all other individuals, and final rankings were assigned based on proximity to this benchmark (shorter distances indicating higher ranks). The Spearman's rank-order correlation between the proposed Mahalanobis-based ranking and the traditional linear method was $\rho = 0.93$, indicating a high degree of agreement between the two systems.

Homepage: https://jem.atu.ac.ir/

Convergent validity was assessed by examining correlations with key concurrent and predictive criteria. The proposed method demonstrated stronger correlations with concurrent criteria (Mathematics score: 0.838; Physics score: 0.756) and a predictive criterion (high school diploma GPA: 0.335) than the traditional method (Mathematics: 0.753; Physics: 0.682; GPA: 0.236).

The overlap between the two methods was 90.22% for the top 10% of candidates and 70.58% for the bottom 10%, suggesting that the proposed method provides better differentiation, particularly among lower-performing individuals. The mean rank change between methods was 10,012.43 (SD = 8,647.66), with a maximum change of 70,472 and a minimum of zero. A key finding was the significantly reduced frequency of tied ranks in the proposed method (145 ties) compared to the traditional method (101,863 ties), demonstrating its superior ability to differentiate between candidates (Siegel & Castellan, 1988). Furthermore, after the removal of outliers, the correlation between the two methods decreased from 0.9343 to 0.8609. In contrast, the internal consistency (intra-method correlation) of the proposed method remained perfect (1.0000), confirming its robustness against extreme values.

Discussion and Conclusion

Accurate candidate ranking in large-scale examinations such as Iran's Konkoor is critical for ensuring equitable university admissions. Traditional methods, which rely on a linear combination of standardized scores, are fundamentally limited by their inability to account for inter-subject correlations and their sensitivity to outliers (Dorans, 2004). This study demonstrates that integrating Principal Component Analysis (PCA) with Promax rotation and Mahalanobis Distance creates a robust framework that significantly enhances ranking precision and differentiation. The high Spearman's rank correlation ($\rho = 0.93$) with the traditional method confirms strong overall compatibility. However, the substantial reduction in tied ranks (145 versus 101,863) and the markedly stronger correlations with key subjects (Mathematics: +11.3%; Physics: +10.8%) underscore the superior discriminatory power and performance of the proposed methodology.

The proposed method mitigates outlier impact by leveraging covariance structures. The post-outlier removal correlation drop of

Homepage: https://jem.atu.ac.ir/

7.8% indicates its differential handling of extreme values. PCA's dimensionality reduction preserved 83.56% of variance, ensuring robust data representation (Jolliffe & Cadima, 2016). These findings align with prior studies (Twenefour et al., 2015; De Maesschalck et al., 2000) and confirm ability of the method to enhance fairness and precision in candidate ranking.

Acknowledgments

Future research should explore the integration of machine learning techniques, such as neural networks, with this methodology and should validate its efficacy across other academic exam groups. Policymakers and testing authorities are encouraged to consider adopting frameworks incorporating Principal Component Analysis and Mahalanobis Distance to improve the fairness and accuracy of large-scale candidate ranking.

Homepage: https://jem.atu.ac.ir/

اندازه‌گیری تربیتی
https://jem.atu.ac.ir/

دوره ۱۷، شماره ۶۳، ۱۴۰۵، ۱-۳۹
ISSN: 2252-004x
eISSN: 2470-6240

مقایسه رتبه‌بندی داوطلبان آزمون سراسری با استفاده از ترکیب خطی نمرات تراز، با روش تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس

مه سیما اربابی دانشجوی دکتری رشته سنجش و اندازه‌گیری، دانشگاه تهران، تهران، ایران. رایانامه: mahsima.arbabi123@gmail.com

ابراهیم خدایی* *نویسنده مسئول*، استاد تمام، گروه روش‌ها و برنامه‌های آموزشی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران، تهران، ایران. رایانامه: khodaie@ut.ac.ir

کیوان صالحی دانشیار، گروه روش‌ها و برنامه‌های آموزشی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران، تهران، ایران. رایانامه: keyvansalehi@ut.ac.ir

چکیده

رتبه‌بندی داوطلبان آزمون سراسری یکی از مراحل کلیدی در فرآیند پذیرش دانشگاهی است. روش‌های سنتی رتبه‌بندی که بر پایه ترکیب خطی نمرات تراز استوارند، با محدودیت‌هایی نظیر عدم توجه به توزیع داده‌ها و همبستگی میان متغیرها مواجه هستند. این پژوهش با هدف ارتقای فرآیند رتبه‌بندی، از تحلیل مؤلفه‌های اصلی (PCA) برای کاهش ابعاد داده‌ها و از فاصله ماهالانوبیس برای تعیین رتبه داوطلبان بهره برده است. جامعه آماری این مطالعه شامل ۱۲۶, ۷۲۸ داوطلبان آزمون سراسری در گروه آزمایشی علوم ریاضی و فنی در سال ۱۴۰۰ بود که پس از حذف داده‌های گمشده در متغیرها، تعداد داده‌های مورد تحلیل به ۱۱۰, ۵۵۰ مورد کاهش یافت. داده‌ها پس از پیش‌پردازش، با استفاده از تحلیل مؤلفه‌های اصلی کاهش بعد داده و سپس با روش فاصله ماهالانوبیس رتبه‌بندی شدند. برای تحلیل داده‌ها از نرم‌افزارهای SPSS و پایتون استفاده شد. نتایج پژوهش نشان داد که روش پیشنهادی در مقایسه با روش سنتی، از دقت بالاتری در تفکیک داوطلبان برخوردار است و توانایی حفظ اطلاعات کلیدی را دارد و درنتیجه، زمینه را برای پذیرش دانش‌آموزانی فراهم کند که ازنظر دانش تخصصی در سطح بالاتری قرار دارند و شایستگی علمی بیشتری برای ورود به دانشگاه‌ها و مؤسسات آموزش عالی دارند. این یافته‌ها حاکی از آن است که تلفیق تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس می‌تواند رویکردی مؤثرتر برای رتبه‌بندی داوطلبان آزمون سراسری ارائه دهد و زمینه‌ساز بهبود روش‌های کنونی شود.

کلیدواژه‌ها: آزمون سراسری، ترکیب خطی نمرات تراز، تحلیل مؤلفه‌های اصلی، رتبه‌بندی چند متغیره، فاصله ماهالانوبیس

استناد به این مقاله: اربابی، مه سیما، خدایی، ابراهیم، و صالحی، کیوان. (۱۴۰۵). مقایسه رتبه‌بندی داوطلبان آزمون سراسری با استفاده از ترکیب خطی نمرات تراز، با روش تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس. *فصلنامه اندازه‌گیری تربیتی*، ۱۷(۶۳)، ۱-۳۹. https://doi.org/10.22054/jem.2025.85317.3608

© ۲۰۱۶ دانشگاه علامه طباطبائی
ناشر: دانشگاه علامه طباطبائی

نوع مقاله: مقاله پژوهشی
تاریخ دریافت: ۱۴۰۴/۰۱/۲۹؛ تاریخ بازنگری: ۱۴۰۴/۰۴/۲۹
تاریخ پذیرش: ۱۴۰۴/۱۰/۲۲؛ تاریخ انتشار: ۱۴۰۵/۰۱/۰۶

Homepage: https://jem.atu.ac.ir/

مقدمه

رتبه‌بندی در داده‌های چندمتغیره همواره یکی از چالش‌های کلیدی در حوزه آموزش و سایر زمینه‌ها بوده است. رتبه‌بندی خطی و میانگین وزنی به‌عنوان رویکردهای رایج برای مقایسه عملکرد افراد مورداستفاده قرار می‌گرفتند. این روش‌ها کاستی‌هایی ازجمله نادیده گرفتن تأثیرات متغیرهای وابسته و عدم در نظر گرفتن همبستگی بین متغیرها دارند (Njoroge et al., 2014). در آزمون‌های تحصیلی، رتبه‌بندی داوطلبان بر پایه مدل‌های همترازسازی نمرات و مقایسه آماری داوطلبان انجام می‌شود (Dorans, 2004).

در آزمون سراسری ایران (کنکور)، محاسبه نمره کل و رتبه داوطلبان بر اساس ترکیب نمرات تراز شده و سوابق تحصیلی انجام می‌شود. نمره خام هر درس با توجه به تعداد پاسخ‌های صحیح و غلط محاسبه‌شده و سپس برای ایجاد امکان مقایسه میان داوطلبان، تراز می‌شود. نمره کل آزمون از ترکیب نمرات تراز شده دروس عمومی و اختصاصی با ضرایب تعیین‌شده برای هر زیرگروه استخراج می‌شود. در صورت مشمولیت، سوابق تحصیلی داوطلب در محاسبه نمره کل لحاظ شده و بیشترین مقدار ممکن از میان نمرات آزمون و ترکیب‌های مختلف سوابق تحصیلی به‌عنوان نمره نهایی ثبت می‌شود. درنهایت، رتبه داوطلبان در هر زیرگروه به‌طور جداگانه محاسبه‌شده و مبنای پذیرش در رشته‌های دانشگاهی قرار می‌گیرد. تأثیر سوابق تحصیلی بسته به نوع دیپلم و گروه آزمایشی، بین ۵٪ تا ۳۰٪ متغیر است (سازمان سنجش آموزش کشور، ۱۴۰۰).

نمره تراز شده به‌عنوان یک شاخص کلیدی در رتبه‌بندی، ازنظر مفهومی شباهتی به فاصله اقلیدسی دارد، زیرا همبستگی میان دروس را در نظر نمی‌گیرد و مقایسه را بر اساس مقیاسی استاندارد انجام می‌دهد. روش سنتی رتبه‌بندی در ایران، با وجود کارایی بالا و پذیرش عمومی، با محدودیت‌هایی مواجه است. ازجمله این محدودیت‌ها می‌توان به نادیده گرفتن ساختار همبستگی بین دروس، حساسیت به توزیع نمرات و تأثیرپذیری زیاد از داده‌های دورافتاده اشاره کرد. در مقابل، روش‌های آماری پیشرفته‌تر مانند تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس با در نظر گرفتن همبستگی بین متغیرها، امکان رتبه‌بندی دقیق‌تری را فراهم می‌کنند (Holland & Wainer, 2012).

برای رفع این محدودیت‌ها، تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس به‌عنوان دو ابزار آماری قدرتمند در تحقیقات آموزشی به کار گرفته می‌شوند. تحلیل مؤلفه‌های اصلی،

با کاهش ابعاد داده‌های چندمتغیره به فضایی با ابعاد کمتر، ضمن حفظ بخش عمده اطلاعات، تفسیر داده‌های پیچیده را تسهیل می‌کند (Jolliffe & Cadima, 2016). از سوی دیگر، فاصله ماهالانوبیس، معیاری برای سنجش فاصله در فضای چندمتغیره است که با در نظر گرفتن ساختار کوواریانس داده‌ها، فاصله یک نقطه از توزیع را محاسبه می‌کند (De Maesschalck et al., 2000). این دو تکنیک در حوزه آموزش برای ارزیابی عملکرد دانش‌آموزان، شناسایی الگوهای عملکرد و بهبود فرآیندهای تصمیم‌گیری کاربرد گسترده‌ای دارند و به ارتقای دقت و کارایی تحلیل‌های آموزشی کمک می‌کنند.

روش سنتی رتبه‌بندی داوطلبان آزمون سراسری، اگرچه پرکاربرد است، اما به دلیل نادیده گرفتن همبستگی دروس و حساسیت به داده‌های دورافتاده، محدودیت‌هایی در دقت و عدالت دارد (Dorans, 2004). تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس، به‌تنهایی یا در ترکیب، ابزارهایی پیشرفته برای رفع این کاستی‌ها ارائه می‌دهند، اما کاربرد آن‌ها در کنکور ایران کمتر بررسی شده است. ازاین‌رو، پژوهش حاضر با دو سؤال هدایت می‌شود: ۱) آیا روش مبتنی بر ترکیب تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس در مقایسه با روش سنتی ترکیب خطی نمرات تراز، دقت و تمایز در رتبه‌بندی داوطلبان آزمون سراسری را بهبود می‌بخشد؟ ۲) چگونه ترکیب این دو روش می‌تواند تأثیر داده‌های دورافتاده را در رتبه‌بندی کاهش داده و ساختارهای همبستگی عملکرد را بهتر نمایان کند؟ هدف این مطالعه، پاسخ به این سؤالات از طریق مقایسه تطبیقی و تحلیل داده‌های آزمون سراسری است.

پیشینه پژوهشی

تحلیل مؤلفه‌های اصلی به‌عنوان ابزاری کلیدی در کاهش ابعاد داده‌ها، در مطالعات آموزشی برای شناسایی متغیرهای مؤثر بر عملکرد و رتبه‌بندی به کار رفته است. Njoroge و همکاران (2014) نشان دادند که تحلیل مؤلفه‌های اصلی می‌تواند متغیرهایی مانند درس زبان انگلیسی را به‌عنوان عامل اصلی موفقیت تحصیلی برجسته کند. این روش، با ساده‌سازی داده‌های پیچیده، ساختارهای پنهان یادگیری را آشکار می‌سازد و به گفته Twenefour و همکاران (2015)، در مقایسه با سیستم‌های سنتی نمره‌دهی مبتنی بر معدل، دقت بیشتری در تحلیل عملکرد ارائه می‌دهد. Nunes و همکاران (2015) کاربرد تحلیل مؤلفه‌های اصلی را در مدیریت آموزشی برای شناسایی خوشه‌های مرتبط با روش‌های تدریس تأیید کردند،

Homepage: https://jem.atu.ac.ir/

درحالی‌که Aboagye and Mensah (2016) از آن برای تمایز معیارهای عملکرد در آموزش ریاضی و آمار بهره بردند. همچنین، Tsoni و همکاران (2021) با ترکیب تحلیل مؤلفه‌های اصلی و تکنیک‌های خوشه‌بندی، الگوهای رفتاری دانش‌آموزان را در آموزش از راه دور بررسی کردند. در حوزه یادگیری الکترونیکی، Raj and Vidyaathulasiraman (2021) و Majid و همکاران (2023) نشان دادند که تحلیل مؤلفه‌های اصلی به‌عنوان ابزار پیش‌پردازش، بهره‌وری یادگیری و دقت پیش‌بینی عملکرد را افزایش می‌دهد. بااین‌حال، این روش همبستگی بین متغیرها را به‌صورت غیرمستقیم کاهش می‌دهد و فاصله واقعی داده‌ها را به‌طور مطلق اندازه‌گیری نمی‌کند.

فاصله ماهالانوبیس، به‌عنوان معیاری قوی در سنجش فواصل آماری در فضای چندمتغیره، برای غلبه بر محدودیت‌های روش‌های سنتی رتبه‌بندی تحصیلی معرفی شده است. De Maesschalck و همکاران (2000) نشان دادند که این روش با در نظر گرفتن ساختار کوواریانس، امکان شناسایی تفاوت‌های ظریف میان گروه‌ها را فراهم می‌کند. در حوزه آموزش، Giovannella و همکاران (2013) کاربرد آن را در تحلیل تأثیر ابزارهای Web 2.0 بر یادگیری دیجیتال تأیید کردند، درحالی‌که Singh (2022) و Kumar and Thakur (2019) از کارایی آن در رتبه‌بندی دانشگاه‌ها بر اساس شاخص‌هایی چون نرخ فارغ‌التحصیلی و تولید علمی گزارش دادند. همچنین، Raj and Vidyaathulasiraman (2021) نشان دادند که این روش در بهینه‌سازی فعالیت‌های یادگیری الکترونیکی مؤثر است. Arai و همکاران (2021) با ترکیب فاصله ماهالانوبیس و رگرسیون چندگانه، دقت پیش‌بینی نمرات زبان‌آموزی را بهبود بخشیدند. در مطالعات اخیر، Ahmed و همکاران (2024) از این روش برای آشکارسازی تغییرات نامحسوس عملکرد دانش‌آموزان در دوران همه‌گیری کووید-۱۹ بهره بردند، و Adhikari (2023) آن را در تحلیل عوامل اجتماعی-تحصیلی در روان‌شناسی آموزشی به کار گرفت. این روش، با توجه به توانایی‌اش در لحاظ کردن همبستگی متغیرها، ابزاری کارآمد برای رتبه‌بندی دقیق‌تر محسوب می‌شود.

ترکیب تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس رویکردی نوین برای افزایش دقت در رتبه‌بندی و تحلیل داده‌های آموزشی ارائه می‌دهد. در این روش، تحلیل مؤلفه‌های اصلی ابتدا ابعاد داده‌ها را کاهش می‌دهد و سپس فاصله ماهالانوبیس، با در نظر گرفتن ساختار کوواریانس، فواصل را در فضای کاهش‌یافته با دقت بیشتری محاسبه می‌کند. Twenefour

Homepage: https://jem.atu.ac.ir/

و همکاران (2015) نشان دادند که استفاده از تحلیل مؤلفه‌های اصلی به‌عنوان گام اولیه، پیچیدگی داده‌ها را کاهش داده و به فاصله ماهالانوبیس امکان می‌دهد تا تفاوت‌های آماری را به‌طور مؤثرتری ارزیابی کند. همچنین، Mehra and Agrawal (2021) گزارش کردند که این ترکیب نه‌تنها دقت مدل‌های پیش‌بینی عملکرد تحصیلی را بهبود می‌بخشد، بلکه شناسایی دانش‌آموزان در معرض افت تحصیلی را تسهیل می‌کند. این رویکرد، با بهره‌گیری از نقاط قوت هر دو روش —کاهش ابعاد توسط تحلیل مؤلفه‌های اصلی و سنجش دقیق فواصل توسط ماهالانوبیس— ابزاری کارآمد برای تحلیل داده‌های چندمتغیره در آموزش فراهم می‌آورد. بااین‌حال، کاربرد این روش ترکیبی در رتبه‌بندی موردبررسی قرار نگرفته است که ضرورت پژوهش حاضر را برجسته می‌کند.

مبانی نظری: روش‌های رتبه‌بندی در ارزیابی‌های آموزشی

رتبه‌بندی داوطلبان و دانش‌آموزان یکی از ارکان اصلی نظام‌های آموزشی و پذیرش دانشگاهی است که به روش‌های متنوعی انجام می‌شود. در این بخش، مبانی نظری روش‌های مختلف رتبه‌بندی، ازجمله رویکردهای سنتی و نوین، موردبررسی قرار می‌گیرد.

۱. برآورد حداکثر درستنمایی در آزمون‌های آنلاین: روش برآورد حداکثر درستنمایی[1] با تخمین توانایی‌های دانش‌آموزان بدون وابستگی به نمرات خام، رویکردی مؤثر در ارزیابی های آنلاین ارائه می‌دهد. این روش به‌ویژه در آزمون‌هایی که سطح دشواری سؤالات از پیش تعیین نشده و بر اساس عملکرد جمعی برآورد می‌شود، پایایی بالایی دارد. به گفته Padmaji و همکاران (2015)، این رویکرد با تمرکز بر مدل‌های آماری، دقت ارزیابی را بهبود می‌بخشد.

۲. رتبه‌بندی مبتنی بر همتایان: در این روش، دانش‌آموزان آثار یکدیگر را ارزیابی می کنند که به تقویت تفکر انتقادی و کاهش بار کاری کادر آموزشی کمک می‌کند. سیستم خودسازگار رتبه‌بندی همتایان، مبتنی بر الگوریتم (روش مراجع و مراکز) HITS،[2] تفاوت شایستگی ارزیابان را در نظر می‌گیرد و با ترکیب کیفیت کار دانش‌آموز و اعتبار ارزیابی همتایان، سنجشی جامع ارائه می‌دهد (Overveld and Verhoeff, 2013).

1. Maximum Likelihood Estimation
2. Hyperlink-Induced Topic Search

Homepage: https://jem.atu.ac.ir/

۳. گرادیان نزولی تصادفی در آزمون‌های بزرگ‌مقیاس: الگوریتم گرادیان نزولی تصادفی[1] برای رتبه‌بندی در آزمون‌های بزرگ‌مقیاس به کار می‌رود. این روش با مدیریت داده‌های پراکنده و چندبعدی، از منظم‌سازی تیخونوف برای جلوگیری از بیش‌برازش استفاده می‌کند و در بهینه‌سازی تلفات رتبه‌بندی زوجی، راه‌حلی مقیاس‌پذیر ارائه می‌دهد (Tas, 2022).

۴. هوش مصنوعی انسان‌محور و پروفایل‌های تعاملی: رویکرد هوش مصنوعی انسان‌محور با تحلیل داده‌های فرآیندی و پاسخ‌های دانش‌آموزان، پروفایل‌های تعاملی ایجاد می‌کند. این پروفایل‌ها با شناسایی الگوهای تعامل و عوامل مؤثر بر عملکرد، امکان مداخلات آموزشی هدفمند را فراهم می‌کنند (Guo et al., 2024).

۵. تکنیک توسعه‌یافته رتبه‌بندی چندصفتی: این تکنیک با استفاده از تابع سودمندی، دانش‌آموزان را بر اساس دانش نظری و مهارت‌های عملی ارزیابی می‌کند. با وزن‌دهی به معیارهای مختلف، این روش سنجشی متعادل و جامع ارائه می‌دهد و در زمینه‌هایی که هر دو بعد اهمیت دارند، کارآمد است (Borissova et al., 2019).

۶. رتبه‌بندی در آزمون سراسری ایران: فرآیند رتبه‌بندی داوطلبان در آزمون سراسری ایران شامل مراحل مختلفی است که از محاسبه نمرات اولیه آغاز می‌شود و به تعیین جایگاه در زیرگروه‌های مختلف ختم می‌گردد. ابتدا، نمره اولیه هر درس بر اساس تعداد پاسخ‌های صحیح، غلط و کل سؤالات هر درس به دست می‌آید. برای اطمینان از مقایسه عادلانه میان داوطلبان در سال‌های مختلف، که ممکن است سطح دشواری سؤالات در آن‌ها متفاوت باشد، این نمرات با استفاده از میانگین و پراکندگی نمرات کل داوطلبان استانداردسازی می شوند. این فرآیند، که به تراز کردن نمرات معروف است، امکان مقایسه منصفانه را فراهم می‌کند.

در گام بعد، نمره کل داوطلب در هر زیرگروه با در نظر گرفتن نمرات تراز شده دروس عمومی و اختصاصی و وزن‌های خاص هر درس در آن زیرگروه محاسبه می‌شود. برای داوطلبانی که سوابق تحصیلی دارند، این نمره با ترکیب نتایج آزمون و نمرات دیپلم یا پیش‌دانشگاهی (به‌صورت جداگانه یا ترکیبی) تعیین می‌شود و بالاترین مقدار به‌عنوان نمره نهایی انتخاب می‌گردد. درنهایت، رتبه‌بندی در هر زیرگروه بر اساس نمره کل داوطلب و

1. Stochastic Gradient Descent

Homepage: https://jem.atu.ac.ir/

تعداد رقبا در همان زیرگروه انجام می‌شود (سازمان سنجش آموزش کشور، ۱۴۰۰). این روش، که ریشه در مدل‌های آماری دارد، به داوطلبان اجازه می‌دهد از سوابق تحصیلی قوی خود بهره ببرند، اما در صورت تأثیر منفی سوابق، تنها عملکردشان در آزمون ملاک قرار می گیرد.

۷. فاصله ماهالانوبیس در سیستم‌های رتبه‌بندی

فاصله ماهالانوبیس، به‌عنوان معیاری آماری در تحلیل چندمتغیره، فاصله یک نقطه از مرکز توزیع را با در نظر گرفتن ساختار کوواریانس میان متغیرها محاسبه می کند و رویکردی متمایز نسبت به روش‌های سنتی مانند فاصله اقلیدسی ارائه می‌دهد. برخلاف فاصله اقلیدسی، که فرض می کند متغیرها مستقل و واریانس‌ها برابرند، این روش با لحاظ کردن همبستگی ها، تحلیل داده‌های واقعی با روابط پیچیده را ممکن می‌سازد (Roszkowska, 2024).

فاصله ماهالانوبیس در زمینه‌های مختلف آموزشی کاربرد گسترده‌ای یافته است. در سیستم‌هایی مانند روش ماهالانوبیس تاگوچی[1]، از این معیار برای رتبه‌بندی داوطلبان MBA استفاده شده است، جایی که معیارهای چندگانه نظیر نمرات تحصیلی، عملکرد در مصاحبه و سوابق حرفه‌ای با دقت بالا ارزیابی می‌شوند (Ketkar and Vaidya, 2014). همچنین، در تحلیل عملکرد تحصیلی دانش‌آموزان، این روش برای شناسایی الگوهای غیرعادی (مانند نقاط دورافتاده در نمرات) و تفکیک گروه‌های مختلف بر اساس ویژگی‌های چندمتغیره به کار رفته است (Lim and Lanckriet, 2014). در آموزش الکترونیکی، از آن برای بهینه‌سازی سیستم‌های توصیه‌گر و تحلیل رفتار یادگیری استفاده‌شده که به تشخیص نیازهای خاص دانش‌آموزان کمک می کند (Behrendt et al., 2024). در رتبه‌بندی دانشگاه‌ها نیز، این معیار با در نظر گرفتن شاخص‌هایی مانند نرخ فارغ‌التحصیلی و تولید علمی، دقت تحلیل را افزایش داده است (Cheng et al., 2022).

مزایای فاصله ماهالانوبیس

این روش مزایای قابل‌توجهی در رتبه‌بندی آموزشی ارائه می‌دهد. نخست، با بهره‌گیری از ماتریس کوواریانس، چارچوبی مستحکم برای مدیریت همبستگی‌ها و عدم قطعیت فراهم می‌کند و دقت و پایایی را ارتقا می‌بخشد (Cheng et al., 2022). دوم، استفاده از برآوردگرهای قوی، مانند برآوردگرهای انقباض، مقاومت آن را در برابر نقاط دورافتاده

1. MTS

Homepage: https://jem.atu.ac.ir/

افزایش می‌دهد و کارایی را در شناسایی الگوهای غیرعادی بهبود می‌بخشد (Lim and Lanckriet, 2014). سوم، در مقایسه با روش‌های مبتنی بر فاصله اقلیدسی، دقت بالاتری در تصمیم‌گیری چندمعیاره دارد، به‌ویژه در شرایطی که متغیرها همبسته‌اند (Li et al., 2015). چهارم، ازنظر محاسباتی برای پردازش مجموعه داده‌های بزرگ کارآمد است و در تحلیل‌های سریع، مانند تشخیص ناهنجاری در سری‌های زمانی، عملکرد مطلوبی دارد (Behrendt et al., 2024). درنهایت، انعطاف‌پذیری بالای آن امکان استفاده در زمینه‌های متنوع، از تصویربرداری پزشکی تا سیستم‌های آموزشی، را فراهم کرده و قابلیت تعمیم آن را تأیید می‌کند.

چالش‌های فاصله ماهالانوبیس

با وجود این مزایا، پیاده‌سازی فاصله ماهالانوبیس چالش‌هایی به همراه دارد. این روش به بررسی دقیق تنوع داده‌ها و انتخاب دقیق معیارها وابسته است، زیرا انتخاب نادرست یا عدم بهینه‌سازی می‌تواند عملکرد را تضعیف کند (Lim and Lanckriet, 2014). پیچیدگی محاسباتی آن نیز در مجموعه داده‌های بزرگ نیازمند الگوریتم‌های کارآمد است که ممکن است منابع محاسباتی قابل‌توجهی طلب کند. علاوه بر این، فرض نرمال بودن چندمتغیره در برخی داده‌های آموزشی ممکن است برقرار نباشد، که می‌تواند نتایج را تحت تأثیر قرار دهد. اعتبارسنجی جامع و بهینه‌سازی دقیق برای تضمین پایداری و قابلیت اطمینان ضروری است، به‌ویژه در شرایطی که داده‌ها ناهمگن یا پرسروصدا هستند.

روش

این پژوهش از نوع توصیفی-تحلیلی است و جامعه آماری آن شامل تمامی داوطلبان گروه آزمایشی علوم ریاضی و فنی شرکت‌کننده در آزمون سراسری سال ۱۴۰۰ به تعداد ۱۲۶،۷۲۸ نفر است. داده‌های مورداستفاده شامل نمرات استانداردشده (تراز شده) هفت درس اصلی آزمون سراسری، یعنی زبان و ادبیات فارسی، زبان عربی، فرهنگ و معارف اسلامی، زبان خارجی، ریاضیات، فیزیک، و شیمی، به همراه سوابق تحصیلی دیپلم داوطلبان است که از یک مجموعه داده واقعی، ارائه‌شده توسط سازمان سنجش آموزش کشور، استخراج شدند. پس از پیش‌پردازش و حذف داده‌های گمشده در متغیرها، تعداد نمونه‌های قابل‌تحلیل به ۱۱۰،۵۵۰ نفر کاهش یافت. این حذف به‌منظور اطمینان از کامل بودن داده‌ها و افزایش دقت تحلیل انجام شد.

Homepage: https://jem.atu.ac.ir/

روش تحلیل و رتبه‌بندی

برای رتبه‌بندی داوطلبان، از تحلیل مؤلفه‌های اصلی با چرخش پرومکس استفاده شد تا ابعاد داده‌های چندمتغیره کاهش یابد. این روش، نمرات تراز شده دروس و سوابق تحصیلی دیپلم را به دو مؤلفه اصلی تبدیل کرد که بیشترین واریانس داده‌ها را حفظ می‌کنند. چرخش پرومکس با هدف بهبود تفسیرپذیری و در نظر گرفتن همبستگی احتمالی میان مؤلفه‌ها اعمال گردید. پس از استخراج این دو مؤلفه برای هر داوطلب، مقادیر آن‌ها جمع شدند تا یک شاخص ترکیبی به دست آید. داوطلبی که مجموع مؤلفه‌هایش بیشینه بود، به‌عنوان بهترین فرد انتخاب شد، زیرا این حداکثر نشان‌دهنده عملکرد برتر در هر دو مؤلفه بود.

سپس، فاصله ماهالانوبیس هر داوطلب از این فرد برتر در فضای دوبعدی تعریف‌شده توسط مؤلفه‌ها محاسبه شد. این معیار، با لحاظ کردن ساختار کوواریانس میان مؤلفه‌ها، فاصله هر نفر از نقطه مرجع (بهترین داوطلب) را تعیین کرد. رتبه‌بندی نهایی بر اساس این فواصل انجام شد، به‌گونه‌ای که کوتاه‌ترین فاصله از فرد برتر، بالاترین رتبه، و حداکثر فاصله، پایین‌ترین رتبه را به خود اختصاص داد. این فرآیند، تفکیک دقیق‌تری نسبت به روش‌های سنتی فراهم کرد. تحلیل داده‌ها با استفاده از نرم‌افزار SPSS نسخه ۲۱ و زبان برنامه‌نویسی پایتون نسخه ۳,۹,۷ انجام شد تا دقت محاسبات و انعطاف‌پذیری تحلیل تضمین شود. درنهایت، روش رتبه‌بندی سنتی (بر پایه ترکیب خطی نمرات تراز) با روش پیشنهادی از طریق معیارهای مختلفی، ازجمله همبستگی رتبه‌ها، تغییرات رتبه، و تعداد تساوی‌ها، مقایسه شد تا کارایی و دقت هر روش ارزیابی شود.

رتبه‌بندی در آزمون سراسری ایران

در آزمون سراسری ایران، فرآیند محاسبه نمره کل و رتبه داوطلبان از چندین مرحله تشکیل شده است که شامل محاسبه نمرات خام، نمرات تراز شده، تأثیر سوابق تحصیلی، و درنهایت رتبه‌بندی در زیرگروه‌های مختلف است. نمره خام هر درس بر اساس فرمول زیر محاسبه می‌شود:

$$N=\frac{۳T-G}{۳S}\times ۱۰۰$$

که در آن:

- T تعداد پاسخ‌های صحیح،
- G تعداد پاسخ‌های غلط،

Homepage: https://jem.atu.ac.ir/

- S تعداد کل سؤالات هر درس است

نمرات خام ابتدا به نمرات مقیاس نرمال تبدیل‌شده و سپس تراز می‌شوند. در این فرآیند، در گام نخست، توزیع فراوانی نسبی نمرات محاسبه می‌شود. در گام دوم، با استفاده از توزیع فراوانی نسبی و توزیع تجمعی نمرات، رتبه درصدی هر نمره تعیین می‌گردد. در گام سوم، نمره استاندارد متناظر با رتبه درصدی محاسبه‌شده، از طریق معکوس تابع توزیع تجمعی نرمال به‌صورت زیر محاسبه می‌شود:

$$\Phi(z)=\frac{\widehat{Q}(y)}{۱۰۰}=\frac{۱}{\sqrt{۲\pi}}\int_{-\infty}^{z} e^{\frac{-\omega^{۲}}{2}}d\omega$$

که در آن:

- $\Phi(z)$ تابع تراکمی نرمال استاندارد،
- ω متغیر انتگرال‌گیری با دامنه $-\infty$ تا z،
- $\hat{Q}(y)$ نشان‌دهنده رتبه‌ی درصدی است

درنهایت نمره‌های مقیاس شده با استفاده از فرمول زیر به نمرات تراز آزمون سراسری با میانگین ۵۰۰۰ و انحراف استاندارد ۲۵۰۰ تبدیل می‌شوند:

$$sc = \sigma(sc)z + \mu(sc)$$

محاسبه نمره کل آزمون و تأثیر سوابق تحصیلی

نمره کل داوطلب بر اساس ترکیبی از نمرات تراز شده دروس عمومی و اختصاصی و ضرایب مربوطه به زیرگروه‌ها محاسبه می‌شود. نمره کل در هر زیرگروه طبق رابطه زیر تعیین می‌شود:

$$N=\sum_{i=۱}^{n}(N_i\times Z_i)$$

که در آن:

- N_i نمره تراز شده درس i،
- Z_i ضریب آن درس در زیرگروه مربوطه است

برای داوطلبانی که مشمول اعمال سوابق تحصیلی هستند، نمره کل بر اساس یکی از چهار روش زیر محاسبه می‌شود:

۱. فقط بر اساس نمره آزمون

Homepage: https://jem.atu.ac.ir/

۲. بر اساس نمره آزمون + سوابق تحصیلی دیپلم
۳. بر اساس نمره آزمون + سوابق تحصیلی پیش‌دانشگاهی
۴. بر اساس نمره آزمون + سوابق تحصیلی دیپلم و پیش‌دانشگاهی

درنهایت، بالاترین مقدار بین این گزینه‌ها به‌عنوان نمره کل نهایی داوطلب انتخاب می شود.

رتبه‌بندی در زیرگروه‌ها

هر گروه آزمایشی دارای چند زیرگروه است که ضرایب دروس اختصاصی در آن‌ها متفاوت است. رتبه داوطلب در هر زیرگروه به‌طور جداگانه محاسبه‌شده و مبنای پذیرش در رشته‌های دانشگاهی قرار می‌گیرد. تأثیر هر درس در هر زیرگروه طبق جدول ضرایب مشخص‌شده و رتبه‌بندی نهایی بر اساس نمره کل داوطلب در هر زیرگروه و تعداد رقبا در همان زیرگروه انجام می‌شود (سازمان سنجش آموزش کشور، ۱۴۰۰).

نحوه محاسبه نمره و رتبه در آزمون سراسری ایران بر پایه مدل‌های آماری پیشرفته و ترکیب اطلاعات آزمونی و تحصیلی انجام می‌شود. تأثیر مثبت سوابق تحصیلی به داوطلبان این امکان را می‌دهد که در صورت داشتن نمرات تحصیلی بالا، از مزایای آن بهره‌مند شوند، درحالی‌که در صورت تأثیر منفی، فقط نمره آزمون در نظر گرفته می‌شود.

روش پیشنهادی

تحلیل مؤلفه‌های اصلی با چرخش پرومکس

در این پژوهش، برای کاهش ابعاد داده‌های پیچیده و چندمتغیره آزمون سراسری، از تحلیل مؤلفه‌های اصلی استفاده شد. این روش با محاسبه ماتریس کوواریانس، که واریانس‌ها و همبستگی‌های خطی میان متغیرها را اندازه‌گیری می‌کند، و تجزیه مقدار ویژه آن، داده‌ها را به مؤلفه‌های اصلی تبدیل کرد که بیشترین واریانس را توضیح می‌دهند (Dien et al., 2007؛ Salazar Cavazos et al., 2023). دقت این فرآیند به کیفیت برآورد ماتریس کوواریانس وابسته بود و محاسبات با دقت بالا انجام شدند تا اطمینان از صحت نتایج حاصل شود.

به‌منظور افزایش تفسیرپذیری مؤلفه‌های استخراج‌شده، چرخش پرومکس اعمال گردید. برخلاف روش‌های متعامد مانند واریمکس، که برای مؤلفه‌های مستقل مناسب‌اند، پرومکس با ارائه چرخش مورب، همبستگی میان مؤلفه‌ها را مجاز می‌داند و برای داده‌های همبسته، مانند نمرات دروس کنکور، کارایی بیشتری دارد (Adachi, 2004؛ Dien, 2010؛ Mishra

Homepage: https://jem.atu.ac.ir/

et al., 2017). هدف از این چرخش، دستیابی به «ساختار ساده» بود، به‌گونه‌ای که هر مؤلفه با تعداد محدودی متغیر ارتباط قوی و با سایر متغیرها ارتباط ضعیف‌تری داشته باشد. فرآیند در دو مرحله اجرا شد: ابتدا مؤلفه‌ها با تحلیل مؤلفه‌های اصلی استخراج شدند و سپس با پرومکس بهینه‌سازی شدند.

در مقایسه با سایر روش‌ها، پرومکس به دلیل انعطاف‌پذیری در مدیریت همبستگی‌ها انتخاب شد. واریمکس، که واریانس بارگذاری‌های مربع را بیشینه می‌کند، در حضور همبستگی کارایی کمتری دارد (Mishra et al., 2017)، و الگوریتم اینفومکس[1] (به حداکثر رساندن اطلاعات) بیشتر برای تحلیل‌های فضایی و مؤلفه‌های مستقل مناسب است (Dien, 2010). اوبلیمین نیز، اگرچه مشابه پرومکس است، اما پرومکس به دلیل تعادل بین کاهش ابعاد و حفظ اطلاعات ترجیح داده شد.

فاصله ماهالانوبیس: تعریف و محاسبه

فاصله ماهالانوبیس معیاری آماری است که فاصله یک نقطه از مرکز یک توزیع چندمتغیره را با در نظر گرفتن ساختار کوواریانس میان متغیرها محاسبه می‌کند. فرمول استاندارد آن به‌صورت زیر است:

$$D = \sqrt{S^{-1}(x-\mu)^T(x-\mu)}$$

که در آن x بردار مشاهدات، μ بردار میانگین، و S ماتریس کوواریانس است (Escobedo and Mendoza, 2015). برخلاف فاصله اقلیدسی، که فرض برابری واریانس‌ها و استقلال متغیرها را دارد، فاصله ماهالانوبیس با تنظیم واریانس هر بعد و در نظر گرفتن همبستگی‌ها، برای داده‌های واقعی با روابط پیچیده مناسب‌تر است (Roszkowska et al., 2024).

ادغام تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس

برای کاهش پیچیدگی محاسباتی و بهبود کارایی، ابتدا تحلیل مؤلفه‌های اصلی با چرخش پرومکس اعمال شد تا ابعاد داده‌ها کاهش یابد و مؤلفه‌های اصلی استخراج شوند. سپس، فاصله ماهالانوبیس بر اساس این مؤلفه‌ها محاسبه گردید. این ترکیب، با ساده‌سازی داده‌ها و حفظ اطلاعات اساسی، محاسبات را تسریع کرد و دقت تحلیل را افزایش داد (Yin, 2023).

1. Infomax

Homepage: https://jem.atu.ac.ir/

به‌منظور رتبه‌بندی بر اساس عملکرد بیشینه، رویکرد نوینی نیز آزمایش شد که در آن به‌جای میانگین (μ) از حداکثر مقدار مجموع دو مؤلفه کاهش بعد یافته (max) استفاده شد:

$$D_{max}=\sqrt{S^{-1}(x\text{-}max)^{T}(x\text{-}max)}$$

ادغام تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس مزایای متعددی به همراه داشت. نخست، کاهش ابعاد توسط تحلیل مؤلفه‌های اصلی امکان پردازش کارآمدتر داده‌های با ابعاد بالا را فراهم کرد، که در روش‌های سنتی به دلیل نیاز به منابع محاسباتی بالا چالش‌برانگیز است (Yin, 2023). دوم، توانایی فاصله ماهالانوبیس در مدیریت همبستگی میان متغیرها، دقت رتبه‌بندی و شناسایی الگوها را نسبت به معیارهای ساده‌تر مانند فاصله اقلیدسی بهبود بخشید (Roszkowska et al., 2024). سوم، استفاده از حداکثر به‌جای میانگین، دیدگاه متفاوتی را در تحلیل داده‌های حاوی نقاط دورافتاده ارائه داد و استحکام روش را در برابر انحرافات خاص تقویت کرد (Etherington, 2021).

روش‌های ارزیابی و مقایسه با روش سنتی

برای انتخاب روش رتبه‌بندی مناسب در این پژوهش، مجموعه‌ای از معیارهای آماری تعریف شد تا دقت و کارایی روش پیشنهادی (ترکیب تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس) با روش سنتی رتبه‌بندی (بر اساس نمرات کل آزمون سراسری ایران) مقایسه شود. این معیارها با هدف ارزیابی انطباق، پایایی، و تغییرات رتبه‌ها در داده‌های واقعی طراحی شدند و بر اساس دو درس کلیدی (ریاضی و فیزیک و معدل دیپلم) به‌عنوان شاخص‌های عملکرد دانش‌آموزان محاسبه شدند. انتخاب این معیارها بر پایه اصول سنجش و اندازه‌گیری و تحلیل داده‌ها استوار است (Crocker and Algina, 2008).

همبستگی رتبه‌بندی پیشنهادی با روش سنتی

جهت سنجش میزان روایی روش پیشنهادی با روش سنتی، ضریب همبستگی رتبه‌ای اسپیرمن محاسبه شد. این معیار، که توانایی روش در حفظ ترتیب رتبه‌ها را نشان می‌دهد، به دلیل ناپارامتریک بودن و حساسیت به روابط ترتیبی برای مقایسه رتبه‌بندی‌ها مناسب است (Siegel and Castellan, 1988).

پس از انجام فرآیند رتبه‌بندی با استفاده از روش پیشنهادی (ترکیب تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس)، روایی همگرا و واگرا به‌عنوان معیارهایی کلیدی برای اعتبارسنجی روش موردبررسی قرار گرفت. روایی همگرا به‌منظور ارزیابی میزان روایی

Homepage: https://jem.atu.ac.ir/

رتبه‌بندی پیشنهادی با سازه‌های مرتبط (مانند نمرات دروس ریاضی، فیزیک، و معدل دیپلم) محاسبه شد تا اطمینان حاصل شود که روش به‌خوبی توانایی‌های مورد انتظار را بازنمایی می کند. به‌طور هم‌زمان، روایی واگرا بررسی شد تا مشخص شود که رتبه‌بندی پیشنهادی به‌طور مناسب از سازه‌های غیر مرتبط تمایز می‌یابد و از همپوشانی ناخواسته جلوگیری می کند. این تحلیل‌ها با استفاده از ضرایب همبستگی (مانند ضریب اسپیرمن) انجام شدند تا ارتباط روش با معیارهای همگرا و عدم ارتباط آن با معیارهای واگرا تأیید شود. این گام، بر اساس اصول اعتبارسنجی در سنجش و اندازه‌گیری، به‌منظور تضمین دقت و اعتبار علمی رتبه‌بندی اجرا شد (Crocker and Algina, 2008).

همبستگی در ۱۰٪ بالایی و پایینی

به‌منظور تحلیل دقیق‌تر عملکرد روش در گروه‌های افراطی (دانش‌آموزان با عملکرد بسیار بالا یا بسیار پایین)، ضریب همبستگی اسپیرمن برای ۱۰٪ بالایی و ۱۰٪ پایینی رتبه‌ها محاسبه شد. این معیار به ارزیابی توانایی روش در تمایز دقیق بین دانش‌آموزان با عملکرد برجسته یا ضعیف کمک می کند، که در زمینه‌هایی مانند تخصیص منابع آموزشی یا شناسایی نیازهای خاص اهمیت دارد (Roszkowska, 2024). مقایسه این همبستگی‌ها با روش سنتی، برتری یا ضعف روش پیشنهادی را در این بازه‌ها مشخص می کند.

میانگین و انحراف معیار تغییرات رتبه

تغییرات رتبه بین روش پیشنهادی و روش سنتی با محاسبه میانگین و انحراف معیار اختلاف رتبه‌ها[1] سنجیده شد. میانگین تغییرات رتبه نشان‌دهنده میزان جابه‌جایی کلی رتبه‌ها و انحراف معیار بیانگر پراکندگی این تغییرات است (Hastie et al., 2009). مقادیر پایین میانگین و انحراف معیار حاکی از پایایی نسبی روش پیشنهادی نسبت به روش سنتی است، درحالی‌که مقادیر بالا می‌تواند نشان‌دهنده تفاوت‌های قابل‌توجه در اولویت‌بندی دانش‌آموزان باشد.

حداکثر و حداقل تغییرات رتبه

حداکثر و حداقل تغییرات رتبه[2] به‌عنوان شاخص‌هایی از دامنه تغییرات در نظر گرفته شدند. این معیارها نشان می‌دهند که روش پیشنهادی تا چه حد می‌تواند رتبه‌های افراطی را

1. Difference in Ranks
2. Maximum and Minimum Rank Differences

Homepage: https://jem.atu.ac.ir/

تغییر دهد، که برای درک تأثیر آن بر دانش‌آموزان با جابه‌جایی‌های بزرگ مفید است (Dien et al., 2007).

روایی و پایایی روش‌های رتبه‌بندی

به‌منظور مقایسه عملکرد دو روش رتبه‌بندی جدولی از شاخص‌های همبستگی تهیه خواهد شد. این جدول شامل چهار شاخص زیر است:

۱. همبستگی بین رتبه‌بندی‌های دو روش قبل از حذف داده‌های دورافتاده: این شاخص میزان توافق اولیه بین دو روش رتبه‌بندی را بدون در نظر گرفتن داده‌های غیرعادی می سنجد.
۲. همبستگی بین رتبه‌بندی‌های دو روش پس از حذف داده‌های دورافتاده: این شاخص تأثیر حذف داده‌های دورافتاده بر توافق بین دو روش رتبه‌بندی را ارزیابی می‌کند.
۳. همبستگی درون‌روشی برای روش سنتی قبل و بعد از حذف داده‌های دورافتاده: این شاخص پایداری و پایایی روش سنتی را در مواجهه با حذف داده‌های دورافتاده بررسی می‌کند.
۴. همبستگی درون‌روشی برای روش پیشنهادی قبل و بعد از حذف داده‌های دورافتاده: این شاخص پایداری و پایایی روش پیشنهادی را در مواجهه با حذف داده‌های دورافتاده ارزیابی می‌کند.

برای محاسبه این شاخص‌ها، از ضریب همبستگی رتبه‌ای اسپیرمن استفاده خواهد شد. این ضریب به دلیل ماهیت ناپارامتریک و توانایی آن در تحلیل روابط ترتیبی، ابزار مناسبی برای مقایسه رتبه‌بندی‌ها محسوب می‌شود.

هدف از این تحلیل، ارزیابی میزان روایی و پایایی دو روش رتبه‌بندی در حضوروغیاب داده‌های دورافتاده است. همبستگی بین روش‌ها نشان‌دهنده درجه سازگاری آن‌ها با یکدیگر خواهد بود، درحالی‌که همبستگی درون‌روشی پایداری هر روش را در برابر تغییرات ناشی از حذف داده‌های غیرعادی می‌سنجد. این بررسی به درک بهتر تأثیر داده‌های دورافتاده بر رتبه‌بندی‌ها کمک کرده و روشی را که دقت و پایداری بیشتری در شرایط مختلف داده‌ای ارائه می‌دهد، مشخص می‌سازد.

Homepage: https://jem.atu.ac.ir/

تعداد تساوی‌ها[1]

تعداد تساوی‌ها یا رتبه‌های برابر[2] در هر روش محاسبه شد تا تمایزگذاری روش‌ها ارزیابی شود. تعداد بالای تساوی‌ها می‌تواند نشان‌دهنده دقت پایین در تفکیک دانش‌آموزان با عملکرد مشابه باشد، درحالی‌که تعداد کم نشان‌دهنده توانایی بالاتر روش در ارائه رتبه‌بندی های متمایز است (Siegel and Castellan, 1988).

یافته‌ها

در این بخش، نتایج حاصل از تحلیل داده‌های آزمون سراسری سال ۱۴۰۰ با استفاده از روش پیشنهادی (ترکیب تحلیل مؤلفه‌های اصلی با چرخش پرومکس و فاصله ماهالانوبیس) و مقایسه آن با روش سنتی رتبه‌بندی ارائه شده است. تحلیل‌ها بر روی داده‌های ۱۱۰،۵۵۰ داوطلب گروه آزمایشی علوم ریاضی و فنی انجام شد و شامل استخراج مؤلفه‌ها، محاسبه فواصل، و ارزیابی رتبه‌بندی بود. خروجی‌ها با استفاده از نرم‌افزار اس پی اس اس[3] نسخه ۲۱ و پایتون[4] نسخه ۳,۹,۷ محاسبه شدند و به‌صورت توصیفی گزارش می‌گردند.

به‌منظور تجزیه‌وتحلیل داده‌ها از دو روش تحلیل مؤلفه‌های اصلی با چرخش پرومکس و فاصله ماهالانوبیس استفاده شد که در ادامه گزارش آن آمده است. قبل از استفاده از روش تحلیل مؤلفه‌های اصلی، برای اطمینان از مناسب بودن داده‌ها، شاخص کفایت نمونه‌برداری[5] و آزمون کرویت بارتلت بررسی شد. مقدار شاخص کفایت نمونه‌برداری، ۰/۹۱۶ به دست آمد که نشان‌دهنده‌ی کفایت نمونه برای تحلیل عاملی است. همچنین، معناداری آزمون کرویت بارتلت مقدار ۳۲۳۵۶۱/۶۵۸ = (۲۸)χ^2 را نشان داد که همبستگی کافی بین متغیرها برای اجرای تحلیل عاملی وجود دارد. پس از تأیید کفایت داده‌ها، تحلیل مؤلفه‌های اصلی برای استخراج مؤلفه‌ها انجام شد. معیار مقادیر ویژه بیش از یک برای تعیین تعداد مؤلفه‌های مناسب استفاده شد. بر این اساس، دو مؤلفه استخراج شد که درمجموع ۸۳/۵۶٪ از واریانس کل را تبیین کردند. با سهم هر مؤلفه به ترتیب، مؤلفه اول: ۶۲/۸۴٪ و مؤلفه دوم: ۲۰/۷۲٪.

1. Ties
2. Number of Ties
3. SPSS
4. Python
5. Kaiser-Meyer-Olkin (KMO)

Homepage: https://jem.atu.ac.ir/

جدول ۱. مقادیر ویژه و درصد واریانس تبیین‌شده برای دو مؤلفه استخراج‌شده

مؤلفه	مقدار ویژه اولیه		مجموع بارهای مربعی		مجموع چرخش بارهای مربعی
	درصد واریانس	درصد تجمعی	درصد واریانس	درصد تجمعی	
Factor 1	٦٢/٨٤	٦٢/٨٤	٦٢/٨٤	٦٢/٨٤	٣٢٢/٣١
Factor 2	٢٠/٧٢	٨٣/٥٦	٢٠/٧٢	٨٣/٥٦	٩٠٥/٥٣

در ادامه با روش چرخش پرومکس، بارهای عاملی محاسبه‌شده که نتایج آن در جدول ۲ ارائه شده است.

جدول ۲. بارهای عاملی متغیرها پس از چرخش پرومکس

مؤلفه	مقدار ویژه اولیه		مجموع بارهای مربعی	
	۱	۲	۱	۲
تراز زبان و ادبیات فارسی با ضریب	٤١٢٢/٥٥٣	١٠٩٠/٧١٥	٠/٤٧٠	٠/١٢٤
تراز زبان عربی با ضریب	٢٠٢٨/٩٣٤	٧١٦/٤٦٩	٠/٤٦٣	٠/١٦٤
تراز فرهنگ و معارف اسلامی با ضریب	٣٣٧٩/٥٥٧	٨٠٧/٧٢٩	٠/٥١٤	٠/١٢٣
تراز زبان خارجی با ضریب	١٤٧٣/٤٧٥	٦٣٥/٧٢٨	٠/٣٣٨	٠/١٤٦
تراز ریاضیات با ضریب	١٧٧٢/٦٨٧	٢٥٧٠٧/٨٩٧	٠/٠٦٦	٠/٩٦٤
تراز فیزیک با ضریب	٢١٠٧٥/٠٦١	-٣٤٣٩/١٣٨	١/٠٤٩	-٠/١٧١
تراز شیمی با ضریب	٧٩٣٧/٥٢٩	١٢٤٦/٣٥٩	٠/٥٩٦	٠/٠٩٤
سوابق تحصیلی نمره دیپلم	٨٩٨/٦٨٩	٤٢٩/٢٣٦	٠/٤٩٢	٠/٢٣٥

با توجه به بارهای عاملی در جدول ۲، دو مؤلفه استخراج‌شده «مهارت‌های عمومی و علوم پایه» و «توانایی ریاضیات و محاسباتی» تفسیر و نام‌گذاری شدند. مقادیر این دو مؤلفه برای هر داوطلب جمع شدند و داوطلبی که بالاترین مجموع را داشت، به‌عنوان بهترین عملکرد و نقطه مرجع (ماکزیمم داده‌ها) انتخاب شد. این فرد، که در نمودار پراکنش دو مؤلفه (شکل ۱) با رنگ قرمز مشخص شده است، بیشینه جمع دو مؤلفه را در میان همه داوطلبان دارا بود. سپس، فاصله ماهالانوبیس هر داوطلب نسبت به این نقطه مرجع در فضای دوبعدی مؤلفه‌ها محاسبه شد و رتبه‌بندی بر اساس این فواصل، از کوتاه‌ترین تا بلندترین، انجام گردید.

برای بررسی توزیع این فاصله‌ها، جدول ۳ حداقل، حداکثر و پراکندگی مقدار فاصله‌های محاسبه‌شده را ارائه می‌دهد. همچنین، هیستوگرام (شکل ۲) توزیع فاصله ماهالانوبیس

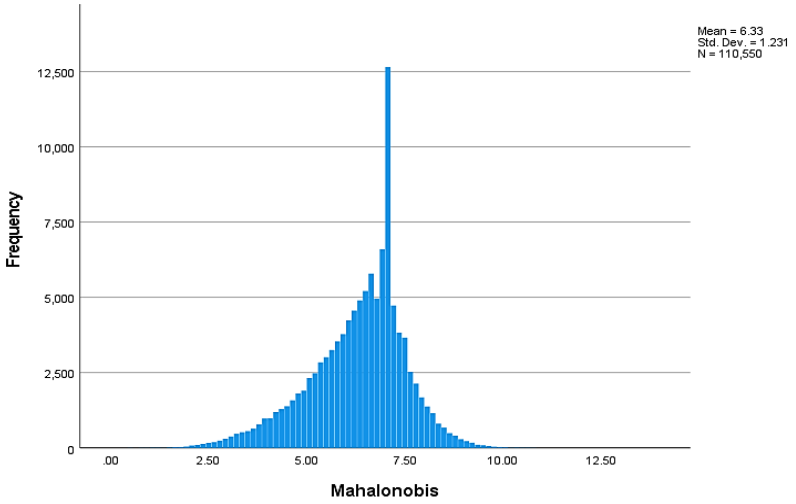
Homepage: https://jem.atu.ac.ir/

داوطلبان را نشان می‌دهد که پراکندگی آن‌ها را نسبت به ماکزیمم داده‌ها در فضای مؤلفه‌های استخراج‌شده به تصویر می‌کشد.

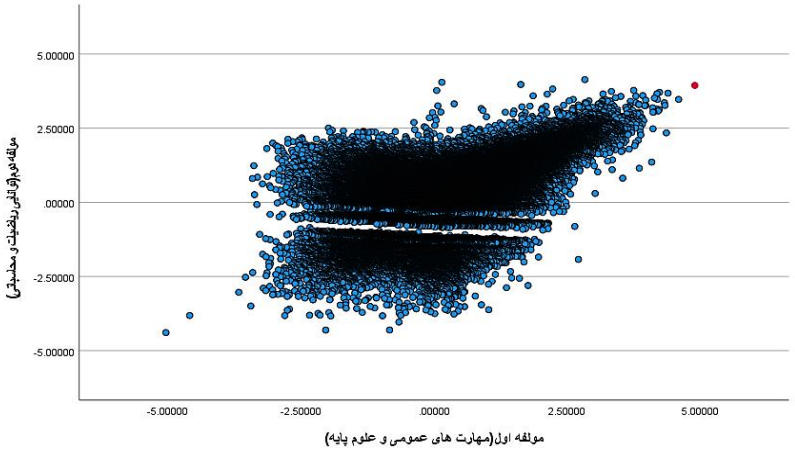
جدول ۳. شاخص‌های توصیفی فاصله ماهالانوبیس در میان داوطلبان کنکور ریاضی

مؤلفه	N	Minimum	Maximum	Mean	Std.Deviation
Factor 1	۱۱۰۵۵۰	-۵/۰٤	٤/۹۰	۰/۰۰	۱/۰۰
Factor 2	۱۱۰۵۵۰	-٤/۳۹	٤/۱٤	۰/۰۰	۱/۰۰
N	۱۱۰۵۵۰				

شکل ۱. نمودار پراکنش و بیشینه مجموع دو مؤلفه



شکل ۲. نمودار هیستوگرام پراکندگی داوطلبان بر اساس فاصله آن‌ها از ماکزیمم توزیع داده‌ها



Homepage: https://jem.atu.ac.ir/

مقایسه نتایج رتبه‌بندی روش پیشنهادی (ترکیب تحلیل مؤلفه‌های اصلی با چرخش پرومکس و فاصله ماهالانوبیس) با روش سنتی (مبتنی بر نمره کل آزمون سراسری) نشان داد که ضریب همبستگی اسپیرمن بین رتبه‌های دو روش ۰/۹۳ است. این مقدار، تطابق بالای رتبه‌بندی پیشنهادی با روش مرسوم را تأیید می‌کند. بااین‌حال روایی همگرا با استفاده از معیارهای ملاک هم‌زمان و پیش‌بین بررسی شد.

معیارهای ملاک هم‌زمان، که بازنمایی هم‌زمان سازه‌های موردسنجش را می‌سنجند، شامل نمرات دروس ریاضی و فیزیک انتخاب شدند. این دروس به دلیل ارتباط مستقیم با توانایی‌های ریاضی و تحلیلی، که هسته اصلی سازه‌های هدف در گروه علوم ریاضی و فنی هستند، معرف‌های مناسبی محسوب می‌شوند. معیار ملاک پیش‌بین، که پیش‌بینی‌کننده عملکرد آینده است، معدل دیپلم در نظر گرفته شد، زیرا این متغیر عملکرد تحصیلی کلی داوطلبان در گذشته را نشان می‌دهد و می‌تواند به‌عنوان شاخصی برای موفقیت آتی عمل کند. ضرایب همبستگی اسپیرمن بین رتبه‌بندی پیشنهادی و این معیارها به ترتیب برای نمرات ریاضی ۰/۸۳۸، برای فیزیک ۰/۷۵۶، و برای معدل دیپلم ۰/۳۳۵ محاسبه شد. این مقادیر بیانگر ارتباط قوی با معیارهای ملاک هم‌زمان و ارتباط متوسط با معیار پیش‌بین است.

در مقایسه، ضرایب همبستگی روش سنتی با همین معیارها به ترتیب ۰/۷۵۳ (ریاضی)، ۰/۶۸۲ (فیزیک)، و ۰/۲۳۶ (معدل دیپلم) بود. روش پیشنهادی در معیارهای ملاک هم‌زمان، یعنی نمرات ریاضی و فیزیک، به ترتیب ۱۱/۳٪ و ۱۰/۸٪ افزایش همبستگی نشان داد، و در معیار پیش‌بین (معدل دیپلم) با بهبود ۴۲/۰٪ عملکرد بهتری داشت. این نتایج حاکی از آن است که روش پیشنهادی توانایی بیشتری در بازنمایی دقیق سازه‌های هم‌زمان (مانند مهارت‌های ریاضی و تحلیلی) دارد و به‌طور نسبی پیش‌بینی‌کننده بهتری برای عملکرد کلی تحصیلی است. این یافته‌ها با چارچوب‌های نظری سنجش و اندازه‌گیری، ازجمله اصول اعتبارسنجی Crocker and Algina (2008)، هم‌خوانی دارد و شواهد محکمی از روایی همگرای روش پیشنهادی ارائه می‌دهد.

Homepage: https://jem.atu.ac.ir/

Homepage: https://jem.atu.ac.ir/

جدول ٤. ضرایب همبستگی روش پیشنهادی و روش سنتی رتبه‌بندی داوطلبان

	نمره خام زبان و ادبیات فارسی	نمره خام زبان عربی	نمره خام فرهنگ و معارف اسلامی	نمره خام زبان خارجی	نمره خام ریاضی	نمره خام فیزیک	نمره خام شیمی	معدل دیپلم	رتبه کشوری	فاصله ماهالانویس
نمره خام زبان و ادبیات فارسی	۱	۰,۵۰۸**	۰,۵۳۷**	۰,۳۱۵**	۰,۳۲۸**	۰,۳۵۸**	۰,۳۸۹**	۰,۱۱۲**	-۰,۶۱۴**	-۰,۴۶۹**
نمره خام زبان عربی		۱	۰,۵۵٤**	۰,۳۲۸**	۰,۳۷۳**	۰,۳۸۸**	۰,۴۳۱**	۰,۱۸۳**	-۰,۶۲۳**	-۰,۵۰۰**
نمره خام زبان خارجی			۱	۰,۳۲٤**	۰,۳۶۲**	۰,۴۱۰**	۰,۴٤٥**	۰,۱۱۲**	-۰,۶٤۷**	-۰,۵۱۸**
نمره خام ریاضی				۱	۰,۲۹۶**	۰,۳۰۶**	۰,۳۰۲**	۰,۱۹٤**	-۰,٤۷۹**	-۰,۳۸۳**
نمره خام زبان خارجی					۱	۰,۳۵۹**	۰,۳۵۸**	۰,۱۷٤**	-۰,۷۵۲**	-۰,۸۳۸**
نمره خام فیزیک						۱	۰,۳۹۱**	۰,۱۷۶**	-۰,۶۸۲**	-۰,۷۵۶**
نمره خام شیمی							۱	۰,۱۸۲**	-۰,۶٤۲**	-۰,۵۵٤**

	فاصله ماهالانوبیس	رتبه کشوری	معدل دیپلم
نمره خام زبان و ادبیات فارسی			
نمره خام زبان عربی			
نمره خام فرهنگ و معارف اسلامی			
نمره خام زبان خارجی			
نمره خام ریاضی			
نمره خام فیزیک			
نمره خام شیمی			
معدل دیپلم	-۰,۳۳۶**	-۰,۲۳۶**	۱
رتبه کشوری	۰,۹۳۷**	۱	
فاصله ماهالانوبیس	۱		

همچنین به‌منظور بررسی میزان تطابق رتبه‌بندی روش پیشنهادی (فاصله ماهالانوبیس نسبت به ماکزیمم) با روش سنتی (نمرات تراز شده)، درصد داوطلبانی که در ۱۰٪ بالای هر دو روش و ۱۰٪ پایین هر دو روش قرار داشتند، محاسبه شد. مقایسه رتبه‌بندی سنتی و رتبه‌بندی مبتنی بر فاصله ماهالانوبیس نشان داد که درصد تطابق رتبه در ۱۰٪ بالایی**: ۹۰/۲۲٪ و درصد تطابق رتبه در ۱۰٪ پایینی**: ۷۰/۵۸٪ است. نتایج ۳,۷,۱. درصد تطابق در ۱۰٪ رتبه‌های برتر نشان داد که ۹۰/۲۲٪ از داوطلبانی که در ۱۰٪ بالای رتبه‌بندی روش سنتی قرار داشتند، در رتبه‌بندی پیشنهادی نیز در این گروه شناسایی شدند. این میزان حاکی از همبستگی قوی بین دو روش در شناسایی داوطلبان برتر است. در مقابل، ۷۰/۵۸٪ از داوطلبانی که در ۱۰٪ پایین روش سنتی قرار داشتند، در روش پیشنهادی نیز در این گروه باقی ماندند. این میزان پایین‌تر از میزان تطابق در رتبه‌های برتر است، که نشان می‌دهد روش پیشنهادی، رتبه‌بندی داوطلبان با عملکرد ضعیف را نسبت به روش سنتی به‌طور متفاوتی انجام داده است. با توجه به تفاوت‌های مشاهده‌شده بین دو روش، مرحله بعدی بررسی میزان جابه‌جایی رتبه‌ها

Homepage: https://jem.atu.ac.ir/

است. در این تحلیل، تغییر رتبه‌های داوطلبان بین دو روش محاسبه‌شده و تأثیر روش پیشنهادی بر دسته‌بندی داوطلبان ارزیابی می‌شود. به‌منظور بررسی میزان تغییر رتبه‌ها بین دو روش رتبه‌بندی، شاخص‌های توصیفی شامل میانگین، انحراف معیار، بیشترین و کمترین تغییر رتبه محاسبه شد. این تحلیل به درک بهتر تفاوت‌های بین روش سنتی (نمرات تراز شده) و روش پیشنهادی (رتبه‌بندی مبتنی بر فاصله ماهالانوبیس نسبت به ماکزیمم) کمک می‌کند. نتایج این تحلیل در جدول ۵ ارائه شده است. نتایج نشان داد که میانگین تغییر رتبه برابر با ۱۰۰۱۲,۴۳ است، که نشان‌دهنده جابه‌جایی قابل‌توجه رتبه‌های داوطلبان بین دو روش رتبه‌بندی است. مقدار انحراف معیار تغییر رتبه ۸۶۴۷,۶۶ به‌دست‌آمده آمد، که نشان می‌دهد تغییر رتبه‌ها در بین داوطلبان به‌طور یکنواخت توزیع نشده است و برخی داوطلبان تغییرات زیادی را تجربه کرده‌اند. بیشترین تغییر رتبه برابر با ۷۰۴۷۲ است، که نشان می‌دهد برخی داوطلبان تغییر رتبه شدیدی را داشته‌اند. در مقابل، کمترین تغییر رتبه برابر با ۰ است، به این معنا که برخی داوطلبان در هر دو روش جایگاه یکسانی داشته‌اند.

جدول ۵. شاخص‌های توصیفی تغییر رتبه بین روش سنتی و روش پیشنهادی

شاخص آماری	مقدار
میانگین تغییر رتبه	۱۰۰۱۲,٤۳
انحراف معیار تغییر رتبه	۸٦٤۷,٦٦
بیشترین تغییر رتبه	۷۰٤۷۲,۰۰
کمترین تغییر رتبه	۰,۰۰

در جدول ۶ تأثیر حذف نقاط دورافتاده بر همبستگی بین روش‌های رتبه‌بندی و پایداری درون‌روشی را نشان می‌دهد.

جدول ٦. تأثیر حذف نقاط دورافتاده بر همبستگی بین روش‌های رتبه‌بندی و پایداری درون‌روشی

شاخص	مقدار
همبستگی بین روش‌های رتبه‌بندی قبل از حذف نقاط دورافتاده	۰,۹۳٤۳
همبستگی بین روش‌های رتبه‌بندی بعد از حذف نقاط دورافتاده	۰,۸٦۰۹
همبستگی درون‌روشی روش سنتی قبل و بعد از حذف نقاط دورافتاده	۱,۰۰۰۰
همبستگی درون‌روشی روش ماهالانوبیس قبل و بعد از حذف نقاط دورافتاده	۱,۰۰۰۰

Homepage: https://jem.atu.ac.ir/

نتایج جدول ۶. نشان‌دهنده عملکرد قابل‌توجه هر دو روش در رتبه‌بندی است. ضریب همبستگی اسپیرمن بین رتبه‌بندی‌های دو روش قبل از حذف نقاط دورافتاده ۰٫۹۳۴۳ به دست آمد که حاکی از ثبات بسیار بالای این دو روش در شرایط اولیه داده‌هاست. بااین‌حال، پس از حذف نقاط دورافتاده، این همبستگی به ۰٫۸۶۰۹ کاهش یافت، که نشان می‌دهد حذف داده‌های دورافتاده تأثیر متوسطی بر توافق بین دو روش داشته است (کاهش ۷٫۸ درصدی). این یافته بیانگر آن است که روش پیشنهادی، اگرچه در کل با روش سنتی هم‌راستاست، اما در مواجهه با داده‌های دورافتاده ممکن است تفاوت‌هایی در اولویت‌بندی ایجاد کند.

از سوی دیگر، همبستگی درون‌روشی برای هر دو روش—روش سنتی و روش پیشنهادی ماهالانوبیس—قبل و بعد از حذف نقاط دورافتاده برابر با ۱/۰۰۰ بود. این مقدار نشان‌دهنده پایایی[1] کامل هر دو روش در برابر تغییرات ناشی از حذف نقاط دورافتاده است و تأیید می‌کند که هر روش به‌طور جداگانه پایداری بالایی در رتبه‌بندی ارائه می‌دهد. بااین‌حال، روایی کامل درون‌روشی در هر دو روش ممکن است نشان‌دهنده حساسیت پایین به نقاط دورافتاده در ساختار محاسباتی آن‌ها باشد، که نیازمند بررسی بیشتر در زمینه تأثیر این داده‌ها بر دقت تمایزگذاری است.

به‌طورکلی، نتایج حاکی از آن است که روش پیشنهادی ماهالانوبیس، ضمن حفظ روایی قابل‌قبول با روش سنتی، پایایی درونی مشابهی را ارائه می‌دهد. کاهش همبستگی بین روش‌ها پس از حذف نقاط دورافتاده می‌تواند به‌عنوان نشانه‌ای از تفاوت در نحوه مدیریت داده‌های غیرعادی توسط روش پیشنهادی تلقی شود، که این امر مزیت بالقوه‌ای در شرایطی با داده‌های ناهمگن محسوب می‌شود.

جدول ۷. روش‌های رتبه‌بندی

روش	تعداد تکرارها
سنتی	۱۰۱۸۶۳
پیشنهادی	۱۴۵

بررسی تعداد تساوی‌ها[2] نشان داد که روش پیشنهادی در مقایسه با روش سنتی عملکرد بهتری در تفکیک و رتبه‌بندی دانش‌آموزان دارد. در روش سنتی، تعداد تساوی‌ها به ۱۰۱۸۶۳

1. Reliability
2. Ties

Homepage: https://jem.atu.ac.ir/

مورد رسید که این رقم بالا حاکی از دقت پایین‌تر این روش در تمایزگذاری بین دانش‌آموزان با عملکرد مشابه است. در مقابل، روش پیشنهادی تنها ۱۴۵ تساوی را ثبت کرد که نشان‌دهنده توانایی قابل‌توجه این روش در ارائه رتبه‌بندی‌های متمایز و دقیق‌تر است. این یافته‌ها با استناد به دیدگاه Siegel and Castellan (1988) تأیید می‌شود که بیان می‌دارد تعداد کم تساوی‌ها به معنای قدرت بالاتر روش در تفکیک افراد است؛ بنابراین، می‌توان نتیجه گرفت که روش پیشنهادی، با کاهش چشمگیر تعداد تساوی‌ها، ابزار مؤثرتری برای ارزیابی و رتبه‌بندی عملکرد دانش‌آموزان فراهم می‌کند و ازاین‌رو، گزینه‌ای مناسب‌تر برای کاربردهای آموزشی محسوب می‌شود.

بحث و نتیجه‌گیری

رتبه‌بندی دقیق داوطلبان در آزمون‌های بزرگ‌مقیاس مانند کنکور سراسری، یکی از چالش‌های اساسی در نظام‌های آموزشی است که مستقیماً بر عدالت و کارایی فرآیند پذیرش دانشگاهی تأثیر می‌گذارد. روش‌های سنتی رتبه‌بندی، که عمدتاً بر نمره‌گذاری استاندارد و ترکیب خطی نمرات تراز استوارند، اگرچه به‌طور گسترده استفاده می‌شوند، اما محدودیت‌هایی نظیر عدم توجه کافی به همبستگی میان متغیرها و حساسیت به داده‌های دورافتاده دارند (Dorans, 2004؛ Kolen and Brennan, 2014). این پژوهش با هدف پاسخ به دو سؤال کلیدی طراحی شد: آیا روش پیشنهادی مبتنی بر ترکیب تحلیل مؤلفه‌های اصلی با چرخش پرومکس و فاصله ماهالانوبیس دقت و تمایز در رتبه‌بندی داوطلبان را بهبود می‌بخشد؟ و چگونه این روش می‌تواند تأثیر داده‌های دورافتاده را کاهش داده و ساختارهای همبستگی عملکرد را بهتر نمایان کند؟ یافته‌های این مطالعه نشان داد که روش پیشنهادی نه‌تنها در هر دو جنبه موفق عمل کرده، بلکه جایگزینی کارآمد برای روش‌های مرسوم ارائه می‌دهد.

پاسخ به سؤال اول: بهبود دقت و تمایز در رتبه‌بندی

یافته‌ها تأیید کردند که روش پیشنهادی دقت و تمایز در رتبه‌بندی داوطلبان را به‌طور قابل‌توجهی بهبود می‌بخشد. ضریب همبستگی اسپیرمن بین رتبه‌بندی پیشنهادی و روش سنتی 0/93 بود، که نشان‌دهنده تطابق نسبی این دو روش است. بااین‌حال، بررسی دقیق‌تر نشان داد که روش پیشنهادی در تفکیک داوطلبان عملکرد بهتری دارد. به‌عنوان‌مثال، 90/22٪ از داوطلبان در ۱۰٪ برتر روش سنتی در روش پیشنهادی نیز در این گروه باقی ماندند،

Homepage: https://jem.atu.ac.ir/

درحالی‌که این میزان برای ۱۰٪ پایین به 70/58٪ کاهش یافت. این تفاوت بیانگر آن است که روش پیشنهادی، به‌ویژه در رتبه‌بندی داوطلبان با عملکرد ضعیف‌تر، تمایز بیشتری ایجاد می‌کند. همچنین، تعداد تساوی‌ها در روش پیشنهادی (۱۴۵ مورد) در مقایسه با روش سنتی (۱۰۱/۸۳۶مورد) به‌طور چشمگیری کمتر بود، که با دیدگاه Siegel and Castellan (1988) هم‌راستاست و نشان‌دهنده قدرت بالاتر این روش در تمایزگذاری دقیق است.

از منظر آماری، میانگین تغییر رتبه ۱۰،۰۱۲,۴۳ و حداکثر تغییر ۷۰/۷۴۲ نشان داد که روش پیشنهادی در برخی موارد جابه‌جایی‌های قابل‌توجهی ایجاد کرده است، درحالی‌که کمترین تغییر (۰) حاکی از پایایی رتبه برخی داوطلبان بود. این تغییرات عمدتاً به دلیل استفاده از تحلیل مؤلفه‌های اصلی برای کاهش ابعاد و استخراج دو مؤلفه کلیدی («عملکرد عمومی و علوم پایه» و «مهارت‌های ریاضی و تحلیلی») و محاسبه فاصله ماهالانوبیس نسبت به حداکثر عملکرد است. برخلاف روش سنتی که صرفاً به میانگین نمرات تراز وابسته است، روش پیشنهادی با در نظر گرفتن ساختار کوواریانس و تمرکز بر عملکرد بیشینه، تفاوت‌های ظریف‌تر میان داوطلبان را منعکس کرد (De Maesschalck et al., 2000). همبستگی بالاتر با دروس اختصاصی (ریاضی ۰/۸۳۸، فیزیک ۰/۷۵۶) در مقایسه با دروس عمومی (فارسی ۰/۴۶۹، معارف ۰/۵۱۸) نیز نشان داد که این روش به‌طور مؤثری هدف پژوهش، یعنی شناسایی داوطلبان قوی در دروس تخصصی، را محقق کرده است؛ بنابراین، پاسخ به سؤال اول مثبت است و روش پیشنهادی به‌عنوان ابزاری دقیق‌تر و متمایزکننده‌تر تأیید می‌شود.

پاسخ به سؤال دوم: کاهش تأثیر داده‌های دورافتاده و نمایان‌سازی همبستگی‌ها

در پاسخ به سؤال دوم، روش پیشنهادی با استفاده از تحلیل مؤلفه‌های اصلی ساختارهای پنهان داده‌ها را آشکار کرد و فاصله ماهالانوبیس با لحاظ کردن کوواریانس، تحلیلی چندمتغیره از فواصل ارائه داد (De Maesschalck et al., 2000). برای بررسی حساسیت به داده‌های دورافتاده، داوطلبانی که فاصله ماهالانوبیس آن‌ها از مقدار بحرانی فراتر بود، حذف شدند. نتایج جدول ۶ نشان داد که همبستگی بین دو روش قبل از حذف پرت‌ها ۰/۹۳۴۳ و پس از حذف ۰/۸۶۰۹ بود، که کاهش ۷/۸ درصدی را نشان می‌دهد. این تغییر حاکی از آن است که روش پیشنهادی در حضور داده‌های دورافتاده، اولویت‌بندی متفاوتی نسبت به روش سنتی ایجاد می‌کند، اما هر دو روش با همبستگی درون‌روشی ۱,۰۰۰۰ پایداری بالایی دارند.

Homepage: https://jem.atu.ac.ir/

این پایداری بیانگر حساسیت پایین هر دو روش به پرت‌هاست، اما روش پیشنهادی با تعدیل بهتر تأثیر این داده‌ها، دقت بیشتری در رتبه‌بندی فراهم کرد (Etherington, 2021).

برخلاف روش سنتی که به دلیل تکیه بر نمرات تراز شده در برابر پرت‌ها آسیب‌پذیرتر است، فاصله ماهالانوبیس با استفاده از کوواریانس، تأثیر داده‌های غیرعادی را کاهش داد و ساختارهای همبستگی میان متغیرها را برجسته‌تر کرد. به‌عنوان‌مثال، همبستگی قوی‌تر با دروس اختصاصی و ضعیف‌تر با دروس عمومی نشان‌دهنده تمرکز روش پیشنهادی بر سازه‌های مرتبط و تمایز از سازه‌های غیر مرتبط بود، که روایی همگرا و واگرا را تأیید می کند (Crocker and Algina, 2008) کاهش ابعاد توسط تحلیل مؤلفه‌های اصلی نیز با حفظ بیشترین اطلاعات در دو مؤلفه، امکان تحلیل عمیق‌تر همبستگی‌ها را فراهم کرد (Jolliffe and Cadima, 2016). این ویژگی‌ها به روش پیشنهادی اجازه داد تا الگوهای عملکردی داوطلبان را با وضوح بیشتری نمایان کند و به سؤال دوم پاسخ مثبت دهد.

از دیدگاه نظری، تحلیل مؤلفه‌های اصلی با تبدیل داده‌ها به مؤلفه‌های اصلی، ترکیب جدیدی از ویژگی‌ها ارائه می‌دهد که اطلاعات کلیدی را در ابعاد کمتر حفظ می‌کند (Jolliffe and Cadima, 2016). فاصله ماهالانوبیس نیز با ارزیابی چندمتغیره، تفاوت‌های ظریف میان داوطلبان را با دقت بیشتری منعکس می‌کند (De Maesschalck et al., 2000). ازنظر تجربی، روش پیشنهادی تساوی‌های گسترده روش سنتی را کاهش داد و با میانگین تغییر رتبه ۱۰،۰۱۲,۴۳ و انحراف معیار ۸،۶۴۷,۶۶، نشان داد که در برخی موارد جابه‌جایی های معنی‌داری ایجاد کرده است. این تغییرات با اصول آماری تمایزگذاری دقیق هم‌راستاست و برتری روش پیشنهادی را در ارزیابی عملکرد تأیید می‌کند (Siegel and Castellan, 1988).

این یافته‌ها نشان می‌دهند که ترکیب تحلیل مؤلفه‌های اصلی و فاصله ماهالانوبیس جایگزینی کارآمد برای روش‌های سنتی است. این روش نه‌تنها دقت و تمایز در رتبه‌بندی را افزایش می‌دهد، بلکه با کاهش تأثیر داده‌های دورافتاده و نمایان‌سازی ساختارهای همبستگی، تحلیل جامع‌تری از عملکرد داوطلبان ارائه می‌کند. در مقایسه با مطالعات گذشته، این پژوهش محدودیت‌های روش‌های سنتی را شناسایی کرد و با معرفی رویکردی نوین، گامی در جهت بهبود فرآیندهای ارزیابی آموزشی برداشت. این برتری به‌ویژه در شناسایی

Homepage: https://jem.atu.ac.ir/

داوطلبان با توانمندی‌های تخصصی، مانند ریاضیات، مشهود است و می‌تواند پذیرش دانشگاهی را هدفمندتر و عادلانه‌تر سازد.

یکی از محدودیت‌های این مطالعه، عدم مقایسه روش پیشنهادی با تکنیک‌های یادگیری ماشین، نظیر شبکه‌های عصبی یا رگرسیون چندمتغیره بود. چنین مقایسه‌ای می‌توانست کارایی روش را در برابر مدل‌های پیشرفته‌تر به‌طور دقیق‌تر ارزیابی کند. همچنین، تمرکز بر گروه علوم ریاضی و فنی ممکن است تعمیم‌پذیری یافته‌ها به سایر گروه‌های آزمایشی را محدود کند.

پیشنهاد برای متولیان و پژوهش‌های آینده

پژوهشگران و مسئولان می‌توانند با استفاده از تحلیل مؤلفه‌های اصلی همراه با چرخش پرومکس، داده‌های آزمون‌های بزرگ‌مقیاس را ساده‌سازی کرده و مؤلفه‌های کلیدی مانند «عملکرد عمومی و علوم پایه» و «مهارت‌های ریاضی و تحلیلی» را استخراج کنند. این روش با حذف اطلاعات زائد، پایه‌ای دقیق‌تر برای رتبه‌بندی داوطلبان فراهم می‌کند. علاوه بر این، محاسبه فاصله ماهالانوبیس به‌عنوان معیاری چندمتغیره می‌تواند جایگزین نمرات تراز شود. این فاصله با در نظر گرفتن همبستگی بین متغیرها، تمایز بین داوطلبان را بهبود بخشیده و به شناسایی توانمندی‌های تخصصی کمک می‌کند.

برای افزایش عدالت در رتبه‌بندی، پیشنهاد می‌شود به‌جای میانگین نمرات، حداکثر مجموع مؤلفه‌ها به‌عنوان معیار رتبه‌بندی در نظر گرفته شود. این رویکرد به‌ویژه در گروه‌هایی که مهارت‌های خاصی مانند ریاضیات اهمیت دارند، باعث شناسایی بهتر داوطلبان با عملکرد برجسته در دروس کلیدی می‌شود. با ترکیب این روش‌ها، فرآیند ارزیابی داوطلبان دقیق‌تر و منصفانه‌تر خواهد بود.

همچنین پیشنهاد می‌شود روایی همگرا و واگرای روش پیشنهادی با معیارهای متنوع‌تر، مانند نمرات عملی، نتایج دانشگاهی، یا آزمون‌های استاندارد دیگر، ارزیابی شود. این تحلیل چندجانبه، روایی روش با اهداف سنجش را تقویت کرده و دقت در شناسایی استعدادها را بهبود می‌بخشد.

پژوهش‌های آینده می‌توانند روش پیشنهادی را با الگوریتم‌های یادگیری ماشین، مانند تقویت گرادیان شدید[1] یا شبکه‌های عصبی عمیق، ترکیب کنند. این تلفیق، با کاهش نویز و

1. Extreme Gradient Boosting (XGBoost)

Homepage: https://jem.atu.ac.ir/

تحلیل پیش‌بینانه عملکرد داوطلبان، دقت رتبه‌بندی را افزایش می‌دهد و امکان شناسایی الگوهای پیچیده‌تر در داده‌ها را فراهم می‌کند.

منابع

سازمان سنجش آموزش کشور. (۱۴۰۰). آشنایی با محتوای کارنامه نتایج علمی آزمون سراسری سال ۱۴۰۰. مرداد ۱۴۰۰.

References

Aboagye, E. A., & Mensah, C. (2016). Principal component analysis of students' academic performance in mathematics and statistics. *American Based Research Journal*, *5*.

Adachi, K. (2004). Oblique Promax rotation applied to the solutions in multiple correspondence analysis. *Behaviormetrika*, *31*, 1-12.

Adhikari, A. (2023). Application of Mahalanobis distance in education and educational psychology: A review. *Innovare Journal of Education*, *11*(4), 5-7.

Ahmed, E. A., Karim, M. R., Banerjee, M., Sen, S., Banu, S., & Warda, W. U. (2024). Higher secondary students' performance in math, English, and other science subjects in pre-COVID 19 and during COVID 19 pandemic: A comparative study using Mahalanobis distance. *Theory and Practice in Language Studies*, *14*(3), 854-865.

Arai, M., Tsubaki, H., & Sagisaka, Y. (2021). Evidence-based statistical evaluation of Japanese L2-learners' proficiency using principal component analysis. In *SHS Web of Conferences* (Vol. 102, p. 01005). EDP Sciences.

Behrendt, F., Bhattacharya, D., Mieling, R., Maack, L., Krüger, J., Opfer, R., & Schlaefer, A. (2024, October). Leveraging the Mahalanobis distance to enhance unsupervised brain MRI anomaly detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 394-404). Springer Nature Switzerland.

Borissova, D., & Keremedchiev, D. (2019). Group decision making in evaluation and ranking of students by extended simple multi-attribute rating technique. *Cybernetics and Information Technologies*, *18*(3), 45-56.

Cheng, J., Yagoubi, N., & Others. (2022). An integrated interval Mahalanobis-Taguchi system for turbine blade quality classification based on vibration data with measurement uncertainties. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *236*(3), 456–470. https://doi.org/10.1177/09544062221076366

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*(1), 1-18.

Dien, J. (2010). Evaluating two-step PCA of ERP data with geomin, infomax, oblimin, promax, and varimax rotations. *Psychophysiology*, *47*(1), 170-183.

Dien, J., et al. (2007). Principal components analysis in educational research. *Psychophysiology*, *44*(5), 785-799.

Dien, J., Khoe, W., & Mangun, G. R. (2007). Evaluation of PCA and ICA of simulated ERPs: Promax vs. Infomax rotations. *Human Brain Mapping*, *28*(8), 742-763.

Homepage: https://jem.atu.ac.ir/

Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement, 28*(4), 227-246.

Escobedo, D., & Mendoza, M. (2015). Análisis de la matriz de covarianza en modelos multivariados. *Revista Mexicana de Estadística, 11*(2), 123–140.

Etherington, T. R. (2021). Mahalanobis distances for ecological niche modelling and outlier detection: Implications of sample size, error, and bias for selecting and parameterising a multivariate location and scatter method. *PeerJ, 9*, Article e11436.

Giovannella, C., Scaccia, F., & Popescu, E. (2013, July). A PCA study of student performance indicators in a Web 2.0-based learning environment. In *2013 IEEE 13th International Conference on Advanced Learning Technologies* (pp. 33-35). IEEE.

Guo, H., Johnson, M., Ercikan, K., Saldivia, L., & Worthington, M. (2024). Large-scale assessments for learning: A human-centred AI approach to contextualizing test performance. *Journal of Learning Analytics, 11*(2), 229-245.

Guo, H., Johnson, M., Saldivia, L., Worthington, M., & Ercikan, K. (2024). Human-centered AI for discovering student engagement profiles on large-scale educational assessments. *Journal of Measurement and Evaluation in Education and Psychology, 15*(Special Issue), 282-301.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Holland, P. W., & Wainer, H. (2012). *Differential item functioning*. Routledge.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), Article 20150202.

Ketkar, M., & Vaidya, O. S. (2014). Evaluating and ranking candidates for MBA program: Mahalanobis Taguchi system approach. *Procedia Economics and Finance, 11*, 654-664.

Kolen, M. J., Brennan, R. L., Kolen, M. J., & Brennan, R. L. (2014). Linking. In *Educational measurement* (pp. 487-536). Springer.

Kumar, A., & Thakur, R. R. (2019). Objectivity in performance ranking of higher education institutions using dynamic data envelopment analysis. *International Journal of Productivity and Performance Management, 68*(4), 774-796.

Li, C., Liu, Q., Liu, J., & Lu, H. (2014). Ordinal distance metric learning for image ranking. *IEEE Transactions on Neural Networks and Learning Systems, 26*(7), 1551-1559.

Lim, D., & Lanckriet, G. (2014, June). Efficient learning of Mahalanobis metrics for ranking. In *International Conference on Machine Learning* (pp. 1980-1988). PMLR.

Majid, N. A. A., Hanawi, S. A., Talib, N. I. M., & Yatim, A. I. A. (2023). Prediction model based on continuous data for student performance using principal component analysis and support vector machine. *TEM Journal*, 1201–1210. https://doi.org/10.18421/tem122-66

Mehra, C., & Agrawal, R. (2021). Design and development of hybrid principal component analysis (HPCA) algorithm for academic performance prediction. *OSF Preprints*. https://doi.org/10.17605/OSF.IO/SZUAD

Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., ... & Laishram, M. (2017). Multivariate statistical data analysis-principal component analysis (PCA). *International Journal of Livestock Research, 7*(5), 60-78.

Homepage: https://jem.atu.ac.ir/

Njoroge, E., Njoroge, G., & Muriithi, D. (2014). Evaluating secondary school examination results: Application of principal component analysis. *Journal of Statistical and Econometric Methods*, *3*(2), 31-46.

Nunes, A. O., da Silva, T. E. V., Mota, J. C. M., de Almeida, A. L. F., & Andriola, W. B. (2015). Developing an instrument for assessment of academic management in engineering courses. *IEEE Latin America Transactions*, *13*(1), 264-271.

Overveld, K. van, & Verhoeff, T. (2013). Self-consistent peer ranking for assessing student work: Dealing with large populations. In *Proceedings of the 5th International Conference on Computer Supported Education (CSEDU 2013)*. https://csedu.scitevents.org

Padmaji, V. G., Chaporkar, P., & Belur, M. N. (2015). A novel student ranking algorithm for off-line examinations. *International Journal for Digital Society*, *6*(1), 1011-1020.

Raj, S. A. P., & Vidyaathulasiraman. (2021). Prioritization of e-learners activities using principal component analysis method. *International Journal of Information Technology*, *13*(6), 2439-2451.

Roszkowska, E. (2024a). Modifying Hellwig's method for multi-criteria decision-making with Mahalanobis distance for addressing asymmetrical relationships. *Symmetry*, *16*(1), Article 77.

Roszkowska, E. (2024b). Mahalanobis distance in multi-criteria decision making. *Journal of Decision Systems*, *33*(2), 123-145.

Roszkowska, E., Filipowicz-Chomko, M., Łyczkowska-Hańćkowiak, A., & Majewska, E. (2024). Extended Hellwig's method utilizing entropy-based weights and Mahalanobis distance: Applications in evaluating sustainable development in the education area. *Entropy*, *26*(3), Article 197.

Salazar Cavazos, J., Fessler, J. A., & Balzano, L. (2023). ALPCAH: Sample-wise heteroscedastic PCA with tail singular value regularization. *arXiv preprint*, arXiv:2307.02745.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences*. McGraw-Hill.

Singh, A. P., Yadav, S. P., & Tyagi, P. (2022). Performance assessment of higher educational institutions in India using data envelopment analysis and re-evaluation of NIRF rankings. *International Journal of System Assurance Engineering and Management*, 1-12.

Tas, E. (2022, October). Large scale ranking using stochastic gradient descent. In *Proceedings of the Bulgarian Academy of Sciences* (Vol. 75, No. 10, pp. 1419-1427).

Thakur, M. (2019). Evaluation of academic institutions using Mahalanobis distance-based approach. *International Journal of Educational Management*, *33*(5), 1056–1070. https://doi.org/10.1108/IJEM-10-2018-0335

Tsoni, R., Panagiotakopoulos, C. T., & Verykios, V. S. (2022). Revealing latent traits in the social behavior of distance learning students. *Education and Information Technologies*, *27*(3), 3529-3565.

Twenefour, F. B. K., Nortey, E. N. N., & Baah, E. M. (2015). Principal component analysis of students academic performance. *International Journal of Business and Social Research*, *5*(2).

Yin, J. (2023). Integration of Promax-rotated PCA and Mahalanobis distance for enhanced computational efficiency in multivariate analysis. *Journal of Computational Statistics*, *38*(4), 1125–1140.

Homepage: https://jem.atu.ac.ir/