Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2, 1-26
2025
Publisher: Learning Gate
DOI: 10.55214/26410230.v7i2.6358

Exploring the interdisciplinary connections between statistics and reader psychology: Insights into reading behavior

Jiexuan Liu[1*]
[1]Librarian of Nanjing Normal University, Nanjing, Jiangsu, China; 79821434@qq.com , 34257@njnu.edu.cn (J.L.).

Abstract: This study explores the borrowing patterns of 3,142 readers who accessed 1,170 statistics books over ten years (2014-2023), resulting in 8,896 book borrowings. To analyze user behaviors, K-means clustering identified three groups of readers based on cumulative borrowing frequency and three groups of books based on their borrowers and borrowings. Correspondence analysis revealed relationships between book and reader types, offering insights into their associations. Factor analysis was employed to identify underlying factors within various book categories, utilizing the Chinese library classification system. A logarithmic transformation was applied to the total borrowings to adhere to the assumptions of normality, facilitating linear regression analysis. Results indicate that the popularity of statistics books in Nanjing Normal University stems from the escalating significance of data literacy, the interdisciplinary integration of statistical methods, and technological advancements in education. Reader Group 3 emerged as the largest engaged cohort, with significant interactions across various book categories. Targeted promotions for Book Groups 1 and 3, aimed at Reader Group 1, could effectively enhance user engagement by reflecting their borrowing preferences. Conversely, a tailored understanding of Reader Group 2's interests can facilitate more focused outreach initiatives. The research also reveals distinct influences on borrowing decisions among different reader types, with varying preferences for interdisciplinary and straightforward scientific content. This study enhances the understanding of borrowing habits while providing methodological insights into applying advanced statistical techniques within library science, emphasizing the importance of targeted engagement strategies.

Keywords: *Behavior, Books, Factor, Psychology, Statistics.*

1. Introduction

Statistics is a discipline that plays a crucial role in understanding and describing populations through random sampling. It also plays a critical role in our increasingly complex and interconnected world, serving as a vital tool for informed decision-making and innovation, particularly in engineering disciplines [1]. Key topics in statistics include regression analysis, hypothesis testing, and the utilization of models that account for various factors such as heteroskedasticity and non-linear effects. These statistical tools facilitate the estimation of population parameters based on sample data, which, while inherently subject to estimation errors, provide valuable insights when applied correctly [2]. In higher education, the overarching goal is to equip students with the knowledge and skills necessary for career success. Understanding the interplay between mathematical modeling and statistics can foster innovation and improve research outcomes, benefiting both fields as they confront their unique challenges and shared objectives [3].

By facilitating critical thinking and problem-solving, statistical methods enable engineers to model and simulate real-world phenomena, ultimately influencing the development of new technologies and improving existing systems. In medical research, statistical methods are essential for designing studies,

* Correspondence: 79821434@qq.com , 34257@njnu.edu.cn

analyzing clinical trials, and interpreting epidemiological data. Biostatistics helps understand treatments' effectiveness and identify risk factors for diseases. Statistical techniques are crucial for spatial analysis and geographic information systems (GIS) in geography. Statistics helps understand demographic trends, assess environmental changes, and explore relationships between geographical factors. Geographers can identify patterns and predict urban development, resource allocation, and environmental impacts by analyzing spatial data. Statistics is vital in analyzing environmental data, assessing pollution levels, and modeling ecological systems. It helps researchers make informed decisions about conservation efforts and resource management. The intersections of statistics with other fields, such as mathematical economics and econometrics, underscore its significance. Econometrics, which combines statistical tests with economic theories, enables the formulation of hypotheses through rigorous mathematical analysis [4]. Econometrics applies statistical techniques to economic data to test hypotheses and forecast future trends. It aids in policy formulation and helps economists understand complex market behaviors.

Furthermore, the applied disciplines of statistics encompass various interdisciplinary applications in social science. Surveys and experiments in sociology and political science rely on statistical methodologies for data collection and analysis. It enables the understanding of societal trends and influences on human behavior. In education, statistics is essential for assessing student performance, evaluating educational programs, and researching teaching methodologies. Data analysis helps educators identify learning outcomes and understand the effectiveness of different instructional strategies, leading to evidence-based decisions for improving educational practices. In psychology, statistical analysis studies behavioral patterns, tests theories, and validates psychological instruments. For instance, understanding the structure of reading psychology can benefit from statistical models to analyze reading behaviors and cognitive processes. Statistics is vital for performance analysis, strategy development, and player evaluation in sports. Statistical methods analyze player statistics, team performance, and match outcomes. Concepts such as advanced metrics, predictive modeling, and analytics are increasingly applied to enhance training, optimize game tactics, and improve overall performance. Coaches and analysts use data-driven insights to make informed decisions, while fans enjoy the statistical breakdown of games and player achievements. The interdisciplinarity of statistics enriches research and practice across these fields, promoting a comprehensive understanding of complex phenomena. As disciplines continue to evolve, the integration of statistical methods will likely expand, providing richer insights and fostering collaboration in areas like sports, mathematics, geography, education, and beyond.

Despite arguments that mathematics and statistics are distinct, their foundations share common elements, such as probability and measure theory, which bridge the two fields and enhance their collaborative potential. The effectiveness of statistical analysis often hinges on the accuracy of sampling distributions, which can vary significantly depending on the underlying data. Estimating the standard error of unbiased estimators is essential for evaluating their reliability, particularly when comparing parameters between distinct populations or subpopulations. In this context, stratified random sampling is frequently employed to ensure that the sample reflects the studied populations' characteristics [5]. Recent contributions from optimal transport theory have further enriched the field, emphasizing the importance of applying these concepts to real-world problems to bolster the credibility of empirical findings [6]. Applying a modeling cycle is essential for researchers to address real-world problems effectively. This cycle emphasizes the importance of recognizing the complexity of scenarios and the necessity for interdisciplinary collaboration to enhance statistical methodologies [3]. The rapid advancements in artificial intelligence (AI), machine learning (ML), and other modern computational methods have transformed the landscape of statistics, making it essential for practitioners to adapt and incorporate these innovations. Understanding mathematical statistics helps navigate uncertainty intrinsic to human cognition and statistical analysis [7]. Thus, statistics not only refines our common sense and intuition but also serves as a formal language for rational thinking, integrating seamlessly with our cognitive processes.

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

Developing tailored statistical education that accommodates varying student strengths and interests is essential to ensure they can effectively contribute to the evolving demands of their respective fields. Unfortunately, many HRD doctoral programs tend to structure their curricula based on the preferences of academic leaders rather than empirical data. Consequently, this disconnection can result in the development of researchers who lack the specialized skills needed to meet future research demands and employer expectations [8]. The integration of diverse cultural perspectives in teaching statistics, as highlighted by the Anthropological Theory of Didactics, enhances the educational experience for students. This approach fosters inclusivity and engagement, bridging the gap between statistical concepts and their application in engineering practices [1]. Further supporting the significance of statistical skills, research suggests they are more correlated with academic success than mathematical skills. Many undergraduate business programs require business statistics (BS) and business mathematics (BM) courses due to their importance in understanding business principles.

Statistical education faces several challenges that can hinder effective learning and understanding among students. One major issue is the disparity between teaching methods, particularly concerning using practice problems versus worked examples. Curricula do not align with the empirical needs or career goals of doctoral students, resulting in a disconnect that could hinder their development as capable researchers [9]. While practice problems allow students to gauge their comprehension and identify challenging aspects of statistical issues, research has indicated that studying worked examples can reduce cognitive load and ultimately enhance performance in the initial stages of learning [10]. This suggests that the lack of direct research comparing these methods in introductory statistics creates a gap in pedagogical strategies that could improve student outcomes. Another significant challenge arises from the complexity of statistical communication. Students are tasked with analyzing datasets and producing various descriptive statistics and visualizations. They are also expected to present their findings in a way that is accessible to those with little statistical knowledge [11]. This requirement emphasizes the need for educators to prioritize communication skills alongside statistical proficiency, which may not always be addressed in conventional curricula. Furthermore, the developmental aspect of statistical learning indicates that statistical reasoning evolves as students mature. Understanding the cognitive architecture of statistical learning highlights that children and adults engage with statistical concepts differently. Empirical studies suggest that infants and young children may adopt broader learning strategies than older learners, which has implications for how educational approaches are structured [12]. If these developmental differences are not considered, educators might miss opportunities to tailor instruction to the cognitive capacities of their students, thus affecting learning efficacy. Lastly, the foundational principles of inferential statistics, such as parameter estimation, standard error, confidence intervals, and hypothesis testing, remain critical yet often underemphasized in many educational contexts. Compared to probability theory, the abundance of textbooks dedicated to inferential statistics indicates a significant reliance on advanced mathematical concepts that may alienate students who struggle with foundational understanding [13]. This reliance on theoretical approaches without practical applications can exacerbate students' difficulty grasping essential statistical concepts. Statistical education faces several challenges that can hinder students' development of essential skills. One significant issue is the lack of emphasis on communication skills within statistics and data science courses. As Hildreth, et al. [14] highlight, written and oral communication is critical for statisticians and data scientists. However, these courses often neglect to teach communication strategies explicitly, which may leave students unequipped to convey their findings to nonstatisticians. This gap is concerning, especially given the importance of translating complex statistical concepts into accessible language for clients and collaborators. The approach to engaging students in reading and understanding statistical literature also presents a challenge. Boyask, et al. [15] notes that early reading development is influenced by relationships, and this influence persists into older age. This suggests that educators must foster deeper connections with students to enhance their interest and engagement in reading statistical content. Students' lack of motivation or engagement with volitional reading may contribute to their difficulties in grasping the intricacies of statistical methods and

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

applications. While teachers play a role in influencing students' reading engagement, the material conditions of schooling can impact this relationship. As Boyask, et al. [15] suggested, a critical literacy lens may reveal that systemic factors within educational environments limit students' engagement with reading. This limitation could further affect how they comprehend and utilize statistical information in practical scenarios. These challenges in statistical education highlight the need for more effective teaching methods, improved communication training, consideration of developmental differences in learning, and a balanced curriculum encompassing theoretical and practical statistics. Addressing these problems requires a more integrated approach to teaching communication alongside statistical concepts and an increased focus on fostering relationships that encourage reading and engagement with the material.

The research encounters several significant challenges, primarily in effectively integrating statistics and psychology to create a coherent framework for understanding reading behavior. This integration requires careful alignment of terminologies and methodologies across both fields, compounded by difficulties in assessing how various statistical concepts are perceived by a diverse readership, which influences engagement and comprehension. Additionally, the inherent subjectivity of psychological factors such as motivation, prior knowledge, and apprehension toward statistics complicates the measurement and quantification of these influences, posing challenges in obtaining reliable data. Methodological constraints exist, particularly regarding regression analysis to quantify preferences for statistics literature, which may be limited by data availability, sample size, and potential selection bias. Despite these challenges, the research introduces several innovative elements, including an interdisciplinary framework that elucidates the connections between statistical education and psychological principles, thus enhancing our understanding of reader interaction with complex texts. Furthermore, the application of regression analysis allows for quantifying psychological influences on book preferences, presenting a novel methodological approach to evaluating the dynamics of these preferences over time. The insights generated from this study hold the potential to inform the development of improved educational materials tailored to diverse reading preferences, ultimately fostering greater engagement and comprehension among readers. By applying established psychological theories to statistical learning, this research cultivates a deeper understanding of the cognitive processes involved in assimilating complex statistical information, aiming to enhance pedagogical practices and the reading experience for students and professionals.

2. Previous Research

Statistics is a branch of mathematics, and its principles are deeply rooted in mathematical theories. Statistical methods utilize mathematical concepts such as probability, distributions, and regression analysis to derive insights from data. Mathematics and statistics work together to solve complex modeling problems, optimize processes, and inform decision-making across diverse fields. While mathematics focuses on numerical and geometric problem-solving, statistics is essential for analyzing data and making informed decisions. In today's digital landscape, the exponential growth of data generated by technology and artificial intelligence heightened the importance of data analytics. Equipping learners with robust data analytic skills is crucial as evidence-based decision-making becomes essential. However, fostering positive attitudes towards statistics and data analytics remains a significant challenge for educators [16]. The intersection of statistical methods and advanced analytics is integral to navigating the complexities of modern data science, highlighting the relevance and necessity of a strong foundation in statistics for future generations. The discipline of statistics encompasses the collection, organization, representation, analysis, interpretation, and prediction based on data. It is built upon descriptive statistics and inferential statistics, with mathematical statistics serving as its foundation through probability theory and further mathematical tools such as linear algebra and differential equations [17]. Probability distributions and statistical inference are fundamental to understanding random variables and making predictions about populations based on sample data.

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

Descriptive statistics serve as fundamental analysis techniques utilized to summarize and describe the characteristics of a dataset in a meaningful way. These methods can be categorized into four key aspects. First, they consist of measures that express the central tendency of observations, often referred to as the average position. Second, they capture the dispersion or variability within the data, providing insights into how spread out the data points are. Third, descriptive statistics help describe the frequency distribution's shape or profile, allowing for a clearer understanding of data patterns. Lastly, these statistics can investigate additional unique aspects, such as the association between two characteristics or a series of observations [18]. While descriptive statistics do not enable conclusions beyond the available data, they play a crucial role in interpreting the data [19]. This rudimentary analysis is essential, as statistics is the art of learning from data. Once data is described, analysts can infer characteristics and potentially model observed trends and behaviors. Descriptive statistics provide a means to derive significant insights from data, turning raw information into valuable knowledge that can guide further analysis and decision-making [20]. However, it is important to navigate the dual nature of statistics. On the one hand, statistics can be a potent tool for generating knowledge. However, on the other hand, they can also lead to misleading or false conclusions if not adequately understood and applied [21]. Understanding the proper use of statistics is essential, as it helps differentiate between valid analyses and pseudo-statistics, ultimately emphasizing the importance of recognizing limitations and potential manipulations in statistical data [21].

Econometrics is a vital field that employs statistical methods and quantitative data to test existing economic theories and predict financial trends. Pioneered by figures such as Lawrence Klein, Ragnar Frisch, and Simon Kuznets—who won the Nobel Prize in economics in 1971—econometrics bridges the gap between economic theory and quantitative analysis. While econometricians typically test hypotheses using data, statisticians often build models based on data sets, highlighting the "theory-driven" versus "data-driven" nature of these fields [22]. Dynamic econometrics, for instance, focuses on analyzing chronological data, often dealing with non-stationary economic series that may incorporate trends [23]. It uses models like ARMA, which combines autoregressive and moving average components to create a parsimonious approach to parameter estimation. Recent advancements in spatial econometrics have introduced methods that are especially impactful for analyzing various scientific problems. These include addressing issues related to heteroscedastic disturbances and incorporating prior information into a Bayesian framework. Practical implementation of these advanced models is made accessible through coding in R, STATA, and Python [24]. It forms the backbone of applied econometrics, integrating statistical methods, economic theory, and mathematical principles to analyze and interpret macroeconomic phenomena [25]. This unification enables econometricians to employ statistical models to investigate economic questions and yield insights that inform policy decisions.

Statistics is crucial as both a theoretical framework and a practical tool utilized across various disciplines. Statistical prediction plays a significant role in various fields by providing insights and guiding decisions based on quantitative data. Using statistical intervals, such as confidence and prediction intervals, further illustrates how statistics capture uncertainty. Instead of representing estimates as single values, these intervals provide a range that quantifies potential errors, offering a more comprehensive view of the underlying data [26]. By identifying and eliminating the causes of uncertainty, researchers can find better statistical methods with lower levels of uncertainty, thus enhancing their decision-making processes [27]. This uncertainty evaluation is essential across all scientific disciplines, as statistical methods provide the foundation for data analysis and scientific inquiry. Vital statistics, for instance, encompass essential data about vital events such as births, deaths, marriages, and divorces. These statistics are crucial for demographic analysis, allowing researchers to study population dynamics, birth and death rates, and migration patterns, offering insights into population growth and aging trends [28]. In finance, statistical methods are essential for modeling market returns, volatility, interest rates, and option prices. These models often rely on calibrating unknown distributions and parameters to observed market prices, while also providing reliability measures for both the models and their parameters [29]. Statistics helps describe and understand

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

variability in financial data, which is crucial since successive observations typically do not yield identical results. This framework enables analysts to draw meaningful conclusions from varying data, enhancing decision-making processes. Advanced deep learning techniques have revolutionized protein structure prediction, an essential aspect of the protein folding problem. Highly accurate predictions through deep neural networks facilitate breakthroughs in understanding protein functions and interactions [30]. This underscores the importance of statistical prediction in scientific research and its potential to influence fields like biochemistry. In weather forecasting, researchers harness extensive weather observation data and advanced data science tools to analyze trends and minimize errors across various parameters. This statistical approach enhances accuracy in weather predictions, showcasing how data-driven methodologies can significantly impact decision-making related to environmental factors [31]. In sports, particularly football, predicting match outcomes is complex due to numerous influencing factors like team morale and player performance. Fans and analysts find this predicting process intriguing, though challenging [32]. The application of statistical models, such as Naïve Bayes, random forest, and K-nearest neighbors (KNN), has been explored in predicting football match outcomes through data collected from social media platforms like Twitter. The research found that while traditional classifiers like Naïve Bayes underperformed, models such as random forest and KNN showed improved accuracy by leveraging rich information from social media, indicating the value of contemporary data sources [33]. In educational settings, libraries can utilize statistical prediction methods to analyze borrowing behaviors and optimize resource allocation. By applying algorithms like the Apriori association rule, libraries can predict readers' needs and enhance their services, thus creating a more tailored reading experience [34]. This method demonstrates the broader applicability of statistical prediction in understanding user behavior and improving service delivery. In university libraries, information technology has significantly enhanced the utilization of data processing. A study indicated that over 90% of university teachers and students believe analyzing library entry behavior can improve library management [35]. This demonstrates how statistical tools can help in understanding user interactions and preferences, leading to better resource allocation and improved reader satisfaction. Librarians have long been collecting and disseminating statistics, a practice gaining increased attention from researchers and library professionals. However, there remains a gap in studies exploring the practical applications of statistics within libraries [36].

Artificial Intelligence (AI) and statistics have become increasingly intertwined, highlighting the importance of both fields in today's data-driven world. The growing complexity of AI technologies and statistical methods necessitates a comprehensive understanding of the concepts and the specialized software applications designed to facilitate this learning. Kallivokas [37] emphasizes the significance of deep learning in undergraduate statistics education, suggesting that modern statistical software plays a crucial role in enhancing students' comprehension of statistical concepts and overall statistical literacy. While these applications can implement advanced statistical techniques, their complex environments can be intimidating for students, often resulting in anxiety that hinders learning. Therefore, visualization tools and simulations are vital, as they allow for experimentation and a more engaging discovery of knowledge. On a broader scale, Khandare, et al. [38] note that AI is the foundation for various computer learning applications, impacting numerous fields, including finance, healthcare, and transportation. Their work examines the top libraries for implementing AI concepts across disciplines such as Machine Learning and Data Science. By providing a comparative analysis of these libraries, the authors aim to equip readers with the necessary building blocks for executing AI projects effectively. The transformative role of algorithmic technologies in libraries is discussed by Meesad and Mingkhwan [39] who argue that these technologies are redefining libraries as dynamic, data-driven information centers. Integrating machine learning and data analytics allows libraries to serve their patrons better while posing ethical, social, and political challenges. Librarians must navigate these complexities to maintain core values like privacy and equitable access while leveraging technology for enhanced user engagement. In addition to these perspectives, the emergence of non-traditional data sources presents new challenges for statistical practices. The need for adaptability in statistical systems arises, as non-

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

traditional data—collected for various non-statistical purposes—differs significantly from traditional survey data. Understanding human behavior through automated data collection requires a dual skill set: statistical modeling and the technical capabilities to manage large, complex data sources. This skill set is essential for the burgeoning field of computational social science, highlighting the necessity for statistical authorities to develop competencies in both AI and statistics to adapt to modern data challenges.

While statistics is an indispensable part of higher education, particularly in social sciences, addressing the anxiety surrounding its learning and application is essential to fostering student success and preparing them for future careers. Students' adept in statistical thinking are often better prepared for success in business education [40]. Courses like "Introductory Statistics with R" and "Statistics and Foresight" are designed to familiarize students with quantitative methods, enabling them to analyze socio-economic phenomena and make decisions in uncertain environments [41]. For instance, education in econometrics, such as courses on applied statistics, covers essential topics like regression analysis, hypothesis testing, and time series analysis. This foundation aids in understanding macroeconomic phenomena and informing policy decisions. The essence of econometrics lies in its ability to combine economic theory and statistical principles to estimate parameters and validate results. Ultimately, the success of these models in making predictions underscores the importance of a robust theoretical framework in the field, as the underlying political and economic ideologies will shape the interpretation of the findings. However, these vital courses can evoke significant anxiety among students, a phenomenon often referred to as statistics anxiety. Research has shown that a student's high school background, including major and grades, significantly affects their academic performance in undergraduate statistics programs [42]. Many college students, especially those without a strong mathematics background, perceive statistics as daunting and demanding, which can lead to test anxiety and maladaptive academic behaviors [4]. This anxiety can manifest in various forms, including worry, tension, and physiological symptoms associated with learning and applying statistical concepts [43]. The challenges students face in statistics courses highlight the need for targeted support and pedagogical strategies that address these feelings of anxiety, aiming to improve performance and confidence in statistical learning. Addressing factors contributing to academic performance is vital, especially for challenging majors, such as statistics, which many students avoid due to their perceived difficulty.

The existing research on reading behavior often focuses narrowly on individual fields, such as cognitive science, linguistics, or education, which can limit the scope and depth of findings. Many studies emphasize quantitative data over qualitative insights, ignoring the emotional and psychological dimensions that shape how individuals engage with texts. Moreover, much of the statistical analysis in reading research may not adequately account for the variations in reader experiences, contextual influences, and demographic differences. The significance of exploring the interdisciplinary connections between statistics and reader psychology lies in the potential for a more nuanced understanding of reading behavior. By integrating statistical methods with insights from psychology, researchers can uncover patterns and underlying motivations that influence how people read and comprehend text. Additionally, bridging these disciplines can foster innovative reading instruction and assessment approaches. For instance, leveraging statistical analysis to interpret psychological data could lead to more personalized reading strategies, tailored to individual learning styles and preferences. This interdisciplinary approach also encourages collaboration among researchers from varied fields, promoting a holistic perspective that enriches the overall understanding of reading as a complex cognitive and emotional process, ultimately informing better educational practices and fostering a deeper appreciation for the intricacies of reading.

3. Data and Method

The sample consists of 3,142 readers who borrowed 1,170 statistics books between 2014 and 2023, resulting in 8,896 book borrowings. The research design is shown in Figure 1.

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

3.1. Subject Word Frequency

A comprehensive examination of the 1,170 statistics books was conducted. Data for bibliographic themes was extracted using the CNMARC and UNIMARC 6XX fields. An Excel pivot table was then utilized to generate statistical insights into the frequency of topic-related terms. The resulting word frequency table was imported to create a thematic word cloud representation.

3.2. K-means Cluster

K-means clustering was performed using SPSS to group similar data points. To begin, navigate to Analyze, Classify, and select K-Means Cluster. Specify the desired number of clusters; in this case, three. The optimal number of clusters was determined based on prior knowledge and exploratory analysis. Cluster analysis was conducted for the books based on the total number of readers and borrowing frequency and the 3,142 borrowers based on their total borrowing frequency. Borrowing frequency was accumulated year by year starting from 2014. For instance, the data for 2023 reflects the total borrowing volume from 2014 to 2023, while the data for 2022 includes the total borrowing volume from 2014 to 2022, and so on. If borrowing patterns only capture isolated instances rather than cumulative behaviors over time, it could lead to sample bias. Therefore, cumulative borrowing volumes were used for reader clustering analysis. After reaching convergence, SPSS outputs include the final cluster centers (centroids), the number of cases in each cluster, variance, within-cluster sum of squares, and an optional distance measure between clusters. This output facilitates the interpretation of the formed clusters and provides insights into their characteristics.

3.3. Book-Reader Correspondence

Correspondence Analysis (CA) is a multivariate statistical technique that analyzes relationships between categorical variables. In SPSS, CA visually represents the associations between rows and columns in a contingency table. To perform this analysis, navigate to the Analyze menu, select Dimension Reduction, and then Correspondence Analysis. In the dialog box, move the chosen variables, book and reader types, derived from the cluster analysis to the appropriate boxes. Click on the Statistics button to select additional output options such as test statistics and the contribution of points to the dimensions. Use the Display button to customize the generated plots. SPSS will produce a correspondence map that visually illustrates the relationships among categories; categories closer together on the map are more similar, while those farther apart are more dissimilar.

3.4. Factor Analysis

The researcher collected data from 3,142 readers with loan records for statistics books (C8, O212), totaling 206,893 items borrowed from categories A to Z between 2014 and 2023. The researcher used the Chinese library classification system to identify common factors across various book types. For the factor analysis, the researcher utilized SPSS by navigating to Analyze, then Dimension Reduction, and selecting Factor. The book categories A to Z were designated variables, with values corresponding to book borrowings. Under Description options, the researcher chose the initial solution, coefficients, Kaiser-Meyer-Olkin (KMO) measure, and Bartlett's test of sphericity. Under Extract options, the researcher selected the unrotated factor solution with eigenvalues greater than one, employing the Varimax technique for the rotation method. Additionally, the researcher saved the factor scores for subsequent linear regression analyses.

To launch SPSS AMOS and start a new project select New from the File menu. Use drawing tools to create a structural equation model, employing the rectangle tool for observed variables (measured variables) and the oval tool for latent variables (unmeasured constructs). Double-click on each shape to label the variables appropriately. Utilize the arrow tool to indicate relationships by drawing arrows from independent variables to dependent variables to show direct effects. Draw arrows reflecting mediation or indirect effects, ensuring the direction indicates the hypothesized pathway. Click on Object

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

Properties to set parameter estimates. Save the model and run the analysis by clicking the Calculate Estimates button. After the analysis, check the output for fit indices, path coefficients, and other statistics to assess the model.

3.5. Linear Regression

A logarithmic transformation of readers' total borrowings was conducted to satisfy the assumptions of normality. This transformation was performed in SPSS by selecting the Transform option and choosing Compute Variable. Using the logarithm function, a new column named LogTotal was created, with the original data column (Total) included in the calculation. The researcher utilized the Analyze, Regression, and Linear options in SPSS for the linear regression analysis. The transformed variable, LogTotal, was designated as the dependent variable, while the factor scores obtained from the factor analysis (F1, F2, and F3) served as the independent variables in the regression model. The Enter method was selected, and the reader registration number was used as the case label.

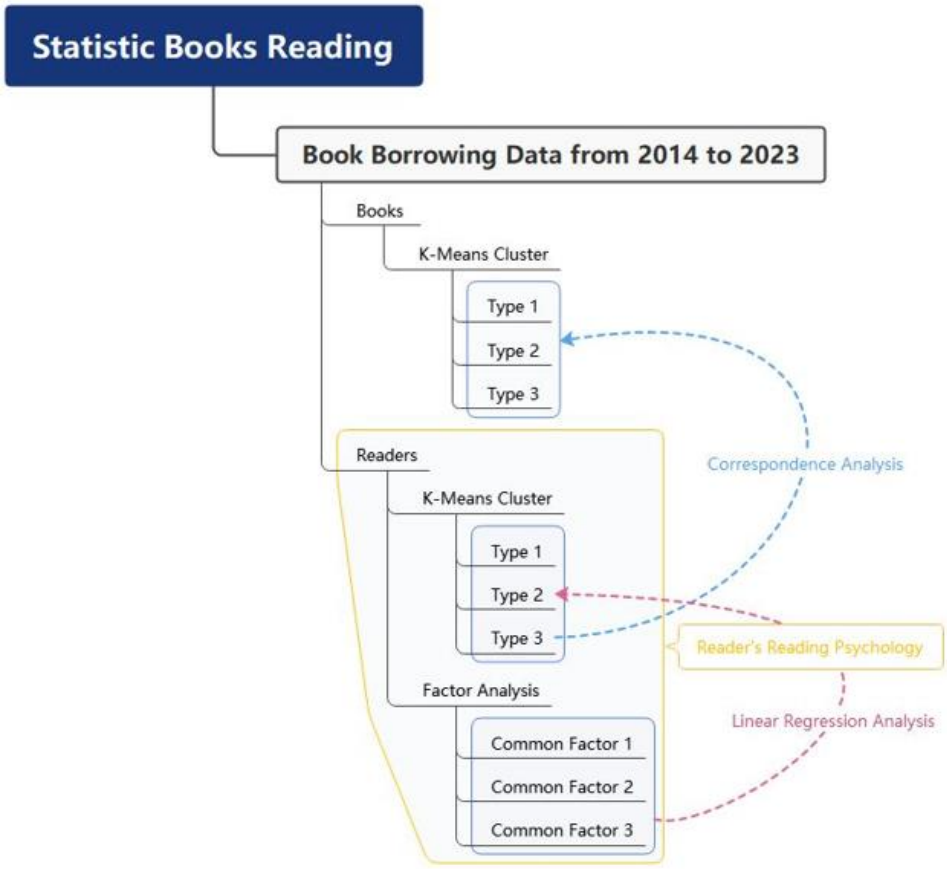

Figure 1.
Research design.

4. Results

4.1. K-Means Cluster of Statistics Books

4.1.1. Popular Subject Words

The library of Nanjing Normal University boasts a comprehensive collection that is intricately aligned with statistical methodologies, software applications, and their practical applications. Figure 2 presents the top 100 popular statistical book keywords in the thematic word cloud map.

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

Statistical Concepts and Techniques: Statistical Analysis (491 occurrences) is the predominant entry, underscoring a robust emphasis on the methodologies inherent to the field of statistics. Additionally, the term Colleges and Universities (324 occurrences) suggests a substantial focus on educational contexts, implying that many resources are tailored for academic instruction and self-study in higher education. The fundamental term Statistics (260 occurrences) further indicates the foundational nature of the subject matter present in the collection.

Software and Tools: The inclusion of terms such as Application Software (161 occurrences), Statistical Program (141 occurrences), and SPSS (78 occurrences) denotes a significant integration of statistical software resources. This trend emphasizes the essential role of computational tools in statistical education and research. Moreover, the mention of Excel (7 occurrences) illustrates its relevance as a widely utilized application for data analysis.

Statistical Methods and Models: The recurring presence of terms like Multivariate Analysis (64 occurrences), Regression Analysis (63 occurrences), Linear Regression (13 occurrences), and Nonparametric Statistics (16 occurrences) indicates a strong focus on a variety of advanced statistical methodologies taught and applied at the institution. The emphasis on Statistical Model (27 occurrences) and Statistical Method (13 occurrences) further illustrates a commitment to understanding the theoretical frameworks that underpin statistical analysis.

Data and Surveys: The prevalence of terms such as Statistical Data (39 occurrences), Sample Survey (25 occurrences), and Statistical Survey (8 occurrences) reflects the significance attributed to data collection methodologies and survey-based research within the academic curriculum. This aspect is critical for fostering a comprehensive understanding of statistical empirical research.

Application of Statistics: The term Applied Statistics (33 occurrences), alongside specific applications such as Medical Statistics (8 occurrences) and Social Statistics (17 occurrences), suggests that the library's collection encompasses materials that apply statistical principles across diverse domains. This variability indicates a curriculum designed to integrate statistical analysis with real-world applications in various fields of study.

Educational Materials: The observation of terms like Teaching Material (12 occurrences) and Graduate Student (37 occurrences) demonstrates a concerted effort to provide resources that support educators and learners in advancing their knowledge of statistics. This alignment signifies an academic environment that fosters statistical literacy at multiple educational levels.

Emerging Trends: Including contemporary terms such as Machine Learning (1 occurrence) and Bayesian Statistics (7 occurrences) reflects the library's responsiveness to emerging methodologies and the evolving landscape of statistical practice and data science.

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

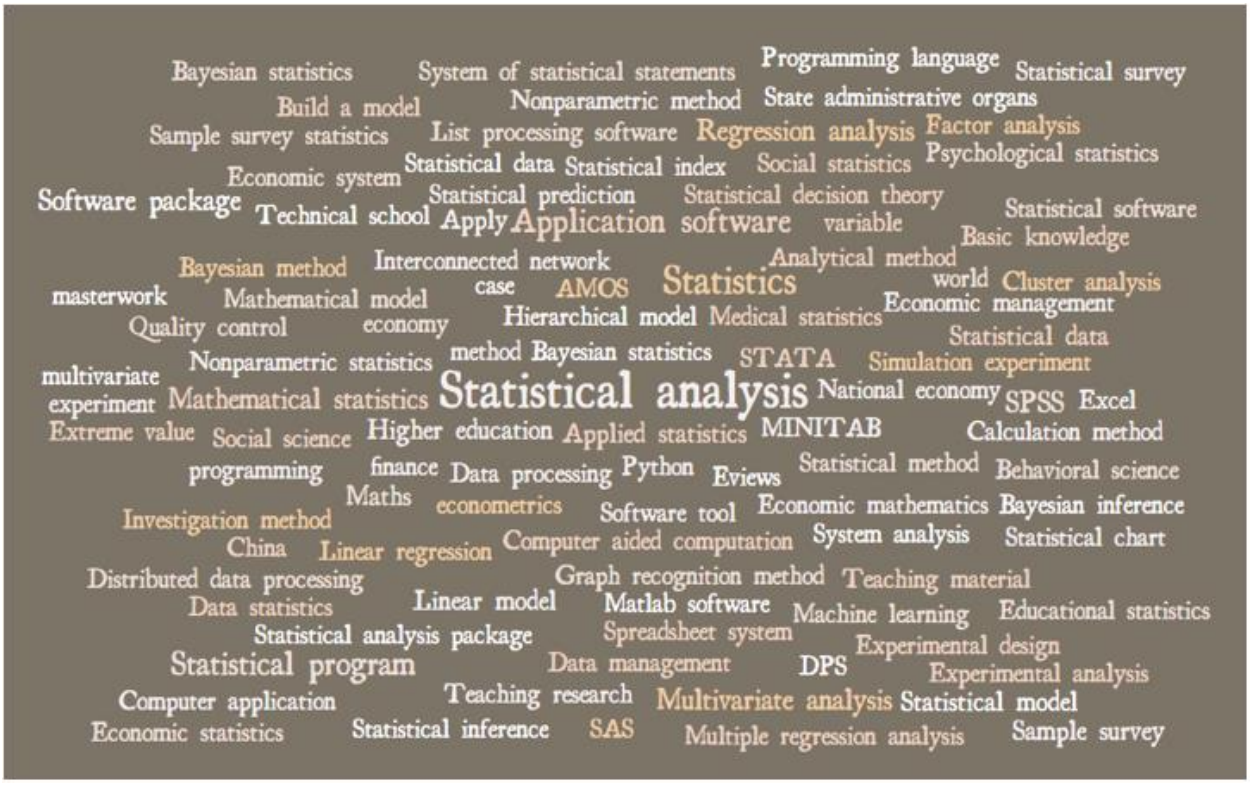

Figure 2.
Subject words.

4.1.2. Three Types of Statistical Books Borrowed

K-means clustering analysis reveals distinct categories among the statistics books based on the features analyzed. Figure 3 presents the clear separation between clusters and the varying number of cases in each, suggesting diverse approaches or topics covered in the statistics literature. This clustering outcome could benefit future decisions regarding these books (such as recommendations or curriculum alignment). The final cluster centers suggest that each cluster has different characteristics based on the features measured, potentially indicating differing styles, complexities, or subjects covered in the books.

The distance between Cluster 1 and Cluster 2 is 79.675, Cluster 1 and Cluster 3 is 33.420, and Cluster 2 and Cluster 3 is 46.266. The relatively large distances, particularly between Cluster 1 and Clusters 2 and 3, hint at significant differences in the features of the books in these clusters. This implies that the books in Cluster 1 are quite different from those in the other clusters. The number of cases in Cluster 1 is 916, Cluster 2 includes 74, and Cluster 3 includes 180. Most books (over 78%) fall into Cluster 1, suggesting that this cluster represents the most common characteristics among the dataset. The large number of books in Cluster 1 can indicate that it captures a widely shared attribute or style prevalent in statistics books. In contrast, Clusters 2 and 3 might represent niche areas with fewer examples.

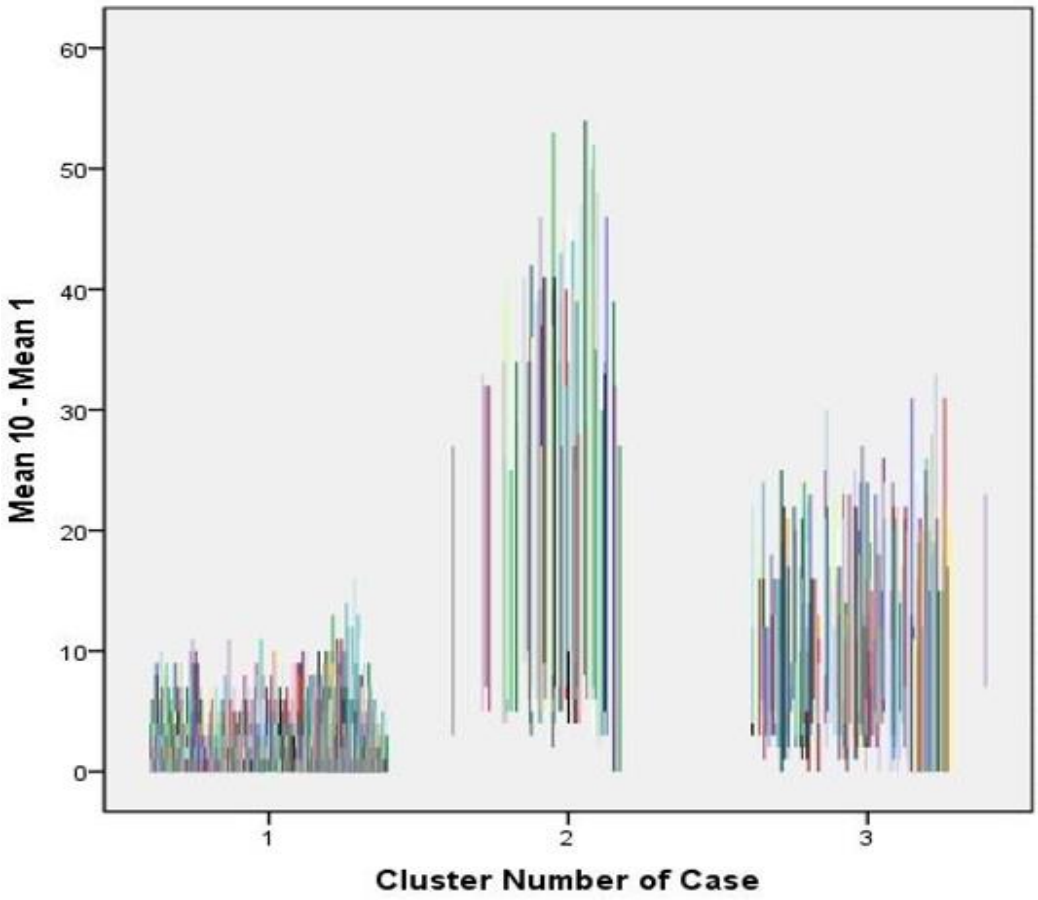

Figure 3.
Clustered high-low-close: Summaries for group cases.

4.2. K-Means Cluster of Book Borrowers

Figure 4 displays the total book borrowings and borrowers in different departments. Cluster 1 has lower values for all parameters, making it likely to represent users who borrow fewer books or engage less with the collection. Cluster 2 displays the highest values, indicating that these borrowers are the most engaged, borrowing significant books. Cluster 3 shows moderate values, suggesting these borrowers are more active, borrowing more books than Cluster 1 but less than those in Cluster 2.

The distance between Clusters 1 and 3 is 39.773, indicating a substantial difference in their profiles. Clusters 1 and 2 have a distance of 16.361, showing that while they are different, they are somewhat closer in terms of their borrowing behavior than Cluster 1 is to Cluster 3. Clusters 2 and 3 have a more considerable distance of 55.683, reflecting a notable distinction in borrowing patterns, with Cluster 3 being significantly more engaged. Cluster 1 contains 303 borrowers, which indicates a smaller, but significant group, likely lower-frequency borrowers. Cluster 2, with 30 borrowers; this cluster is relatively small, suggesting it may represent a niche group with moderate borrowing habits. Cluster 3 represents the majority with 2809 borrowers. This is the largest cluster and suggests a high engagement among users who frequently borrow books.

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

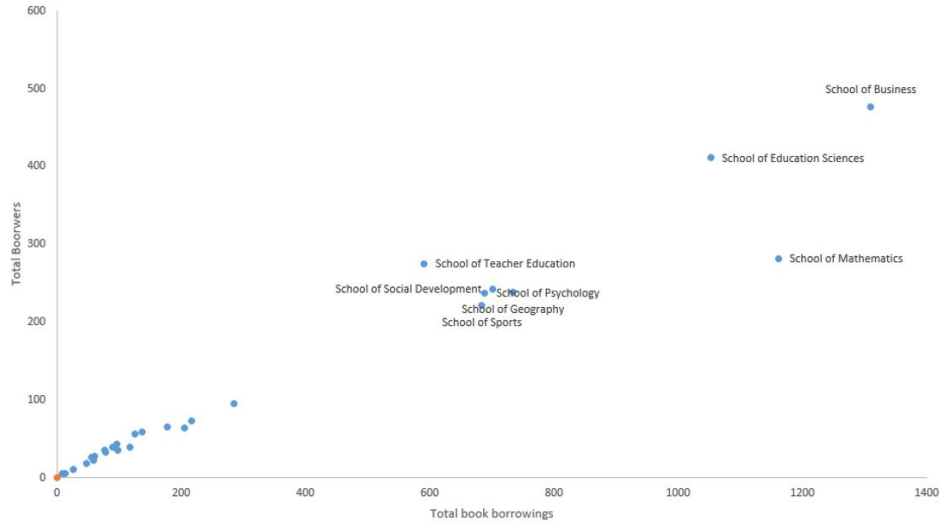

Figure 4.
Total book borrowings and borrowers in different departments.

Furthermore, the total number of book borrowings is 8,896, the total number of borrowers is 3,142. This results in a mean borrowing rate of approximately 2.83 books per borrower. For males, the total borrowings are 2,624, and the total borrowers are 814; the mean borrowings per male borrower is 3.22 books. For females, the total borrowings are 6,272, and the total borrowers are 2,328; the mean borrowings per male borrower is 2.69 books. Males have a higher mean borrowing rate (3.22) than females (2.69), indicating that while there are significantly more female borrowers, male borrowers take out more statistics books on average. Staff members appear to borrow considerably more books on average (6.40) compared to undergraduates (2.58) and postgraduates (2.84). Postgraduates borrow more on average than undergraduates, which might imply a higher need for resources at that level of study. staff members are the highest users of book borrowings relative to the number of staff, indicating possible professional or academic research needs. The data reveals varying borrowing behaviors across different demographics, potentially reflecting their respective needs or engagement levels with the library's resources.

4.3. Book-Reader Correspondence

The correspondence analysis reveals significant relationships between reader preferences and book classifications. By focusing on the strong contributors in the analysis, stakeholders can better understand how to cater to different reader groups' interests and improve engagement with the literature. Figure 5 displays the reader and book group interactions. Reader Group 3 shows considerable engagement across all book groups, particularly Book Group 1, indicating a strong preference or connection.

The Active Margin totals suggest that Reader Group 1 received a total score of 2393 from Book Group 1, while Book Group 3 attracted the highest engagement from Reader Group 3 with a total of 1962. This means Reader Group 3 favors books classified in Book Group 3, reinforcing their specific reading habits. The inertia and singular values reveal that Dimension 1 accounts for a significant proportion of the variance (0.988). This suggests that this single dimension can explain most of the relationship variation. The low values in Dimension 2 indicate less variation or relationship information is captured there. Reader Group 1 contributes significantly to the inertia of Dimension 1 (0.951), suggesting that it plays an essential role in defining the variance among the groups. Reader Group 2 has less impact in Dimension 1 (0.061), suggesting a more specialized or limited engagement with the book groups. Based on these findings, recommendations for personalized reading suggestions or marketing

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

can be established. For instance, if a reader is identified as part of Reader Group 1, targeted promotions for Book Groups 1 and 3 might be effective.

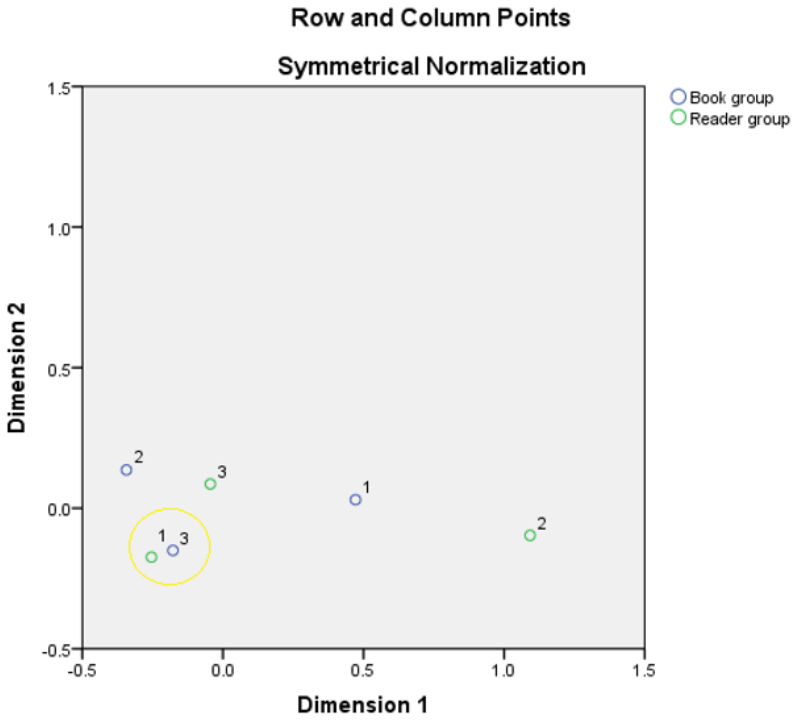

Figure 5.
Reader and book group interactions.

4.4. Common Factors

The Kaiser-Meyer-Olkin value of 0.690 indicates moderate sampling adequacy. A KMO value above 0.60 is generally considered acceptable for conducting factor analysis, suggesting that the data set is suitable for this method. Bartlett's Test of Sphericity (Approx. Chi-Square = 8696.284, df = 253, Sig. = 0.000) shows a significance level. It indicates that the correlation matrix is significantly different from an identity matrix, supporting the idea that underlying factors within the data warrant analysis. The rotated component matrix in Table 1 reveals how each variable loads onto the identified factors. The first factor reflects a blend of social sciences and fundamental natural science. The second factor focuses on socio-political and historical contexts. The third factor is centered on environmental and agricultural disciplines.

Table 1.
Rotated Component Matrix[a].

	Component		
	1	**2**	**3**
C (Sociology)	0.78	0.19	0.06
F (Economics)	0.61	0.08	0.03
N (Theory of Natural Science)	0.50	0.01	0.03
D (Politics)	0.41	0.65	0.01
A (Marxism)	0.13	0.58	-0.01
K (History, Geography)	0.16	0.57	0.11
DF (Law)	-0.11	0.54	-0.01
X (Environment)	0.02	0.00	0.81
S (Agriculture)	0.10	0.01	0.79
P (Astronomy, Earth)	-0.10	0.06	0.53

Note: Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 9 iterations.

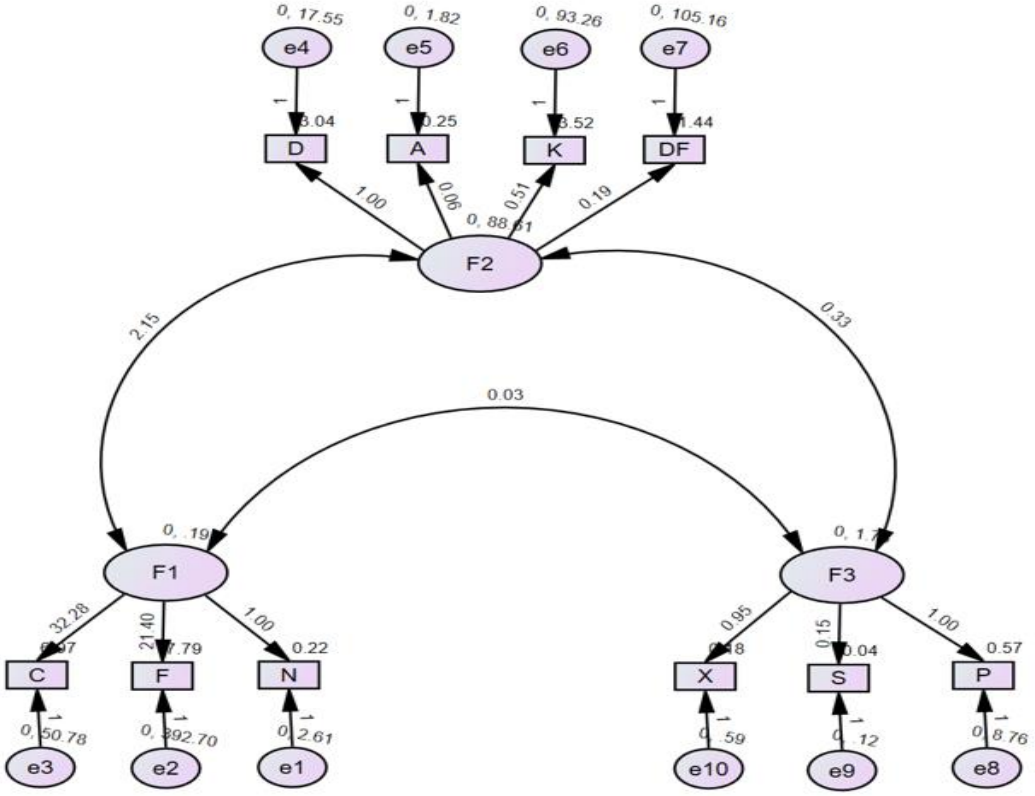

Figure 6.
Structural equation model.

The structural equation modeling (SEM) analysis results provide valuable insights regarding the relationships between the observed variables and the latent factors (F1, F2, F3). C retains a high standardized estimate of 0.892, indicating it substantially affects the latent factor F1. D (0.914) strongly contributes to F2, showing good explanatory power. F (7.788) and C (6.969) have the highest intercept values, which may suggest that these variables have a baseline level that is significantly above zero. The significant covariance between F1 and F2 (2.151, p <0.001) indicates a strong relationship, while F3 and

F2 show a lower non-significant covariance of (0.335, p = 0.231) suggesting weaker interdependence. A high correlation (0.525) between F1 and F2 reiterates their strong relationship. The correlations with F3 are low, hinting that F3 may operate somewhat independently from F1 and F2 in the model. The variance of F2 (88.607) is notably high, indicating that it has a wide variability compared to F1 (0.190) and F3 (1.777), possibly reflecting greater underlying complexity or diversity in the constructs measured by F2. X shows a high squared multiple correlation (0.732), suggesting that a large proportion of its variance is explained by the latent factors. In contrast, N (0.068) indicates that much less of its variance is captured, suggesting it might be less influenced by the model's latent factors (refer to Figure 6).

4.5. Regression Prediction of LogTotal

4.5.1. Type 1 Reader

The R value is 0.590, which suggests a moderate positive correlation between the predictors (F1, F2, and F3) and the dependent variable (LogTotal). The R Square value is 0.348, indicating that the model can explain approximately 34.8% of the variance in LogTotal. This is a modest level of explanatory power. The Adjusted R Square (0.341) is similar and adjusts for the number of predictors in the model, reinforcing that the model explains a significant portion of variability. The F-change statistic of 53.185 indicates that the model is statistically significant, with a p-value (Sig. F Change) of 0.000, which means we can reject the null hypothesis that the model has no predictive power. The Durbin-Watson statistic is 1.576, suggesting this model likely has no significant autocorrelation issue. Figure 7 displays the histogram and standard P-P plot of the regression standardized residual. The ANOVA table shows a significant F-statistic (53.185) with a p-value of 0.000, indicating that at least one of the predictors is significantly related to LogTotal.

The coefficients in Table 2 show the unstandardized and standardized coefficients for each predictor. The unstandardized coefficient for the constant is 1.708, which is the predicted value of LogTotal when all predictors are zero. For the predictor F1, the unstandardized coefficient is 0.201, indicating that for each unit increase in F1, LogTotal is expected to increase by 0.201 units, holding other variables constant. It has a significant t-value of 11.607 ($p < 0.001$), making it an important predictor. The unstandardized coefficient for F2 is 0.038, which is not statistically significant ($p = 0.118$), indicating that changes in F2 do not have a meaningful impact on LogTotal. The unstandardized coefficient for F3 is 0.070, also with a significant t-value of 2.622 ($p = 0.009$), indicating a positive relationship and significance. The Variance Inflation Factor (VIF) values for all predictors are below 2, F1 is 1.071, F2 is 1.073, and F3 is 1.016, indicating that multicollinearity is not a concern. Tolerance values are also acceptable, with all being above 0.1, suggesting that the predictors included in the model are not too highly correlated.

The regression model indicates that F1 and F3 are significant predictors of LogTotal, while F2 does not contribute significantly to the prediction. The model explains a moderate amount of variance in the dependent variable, and there are no major issues with multicollinearity or autocorrelation.

Table 2.
Coefficientsa in 1 type reader prediction.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	1.708	0.020		85.882	0.000					
	F1	0.201	0.017	0.561	11.607	0.000	0.570	0.557	0.542	0.934	1.071
	F2	0.038	0.024	0.076	1.569	0.118	0.221	0.090	0.073	0.932	1.073
	F3	0.070	0.027	0.123	2.622	0.009	0.091	0.150	0.122	0.984	1.016

Note: a. Dependent Variable: LogTotal.

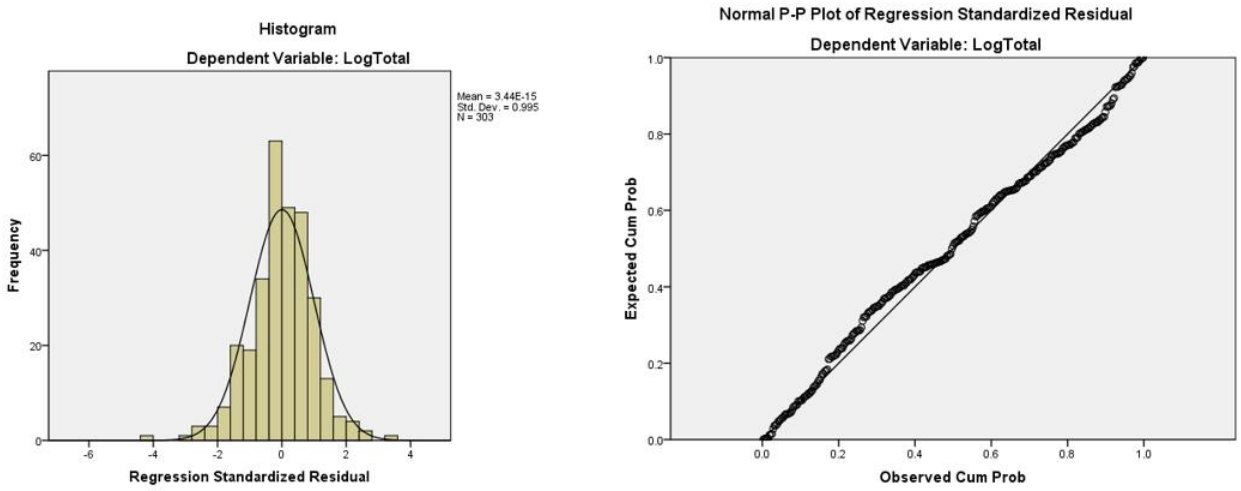

Figure 7.
Histogram and standard P-P plot in type 1 reader prediction.

4.5.2. Type 2 Reader

The regression analysis indicates that the model has a relatively strong relationship with the dependent variable (LogTotal), as evidenced by an R value of 0.728, which translates to an R Square of 0.530. This means that approximately 53% of the variability in LogTotal can be explained by the independent variables (F1, F2, and F3). The Adjusted R Square value of 0.475 suggests that the model is a decent fit after adjusting for the number of predictors. The standard error of the estimate is reported as 31221, and the Durbin-Watson statistic is 1.763, suggesting no autocorrelation in the residuals. Figure 8 displays the histogram and standard P-P plot of the regression standardized residual. The ANOVA table shows a significant F statistic of 9.756, with a p-value (Sig.) 0.000. This indicates that the overall regression model is statistically significant, meaning that at least one of the predictors significantly contributes to explaining the variance in LogTotal.

The coefficients in Table 3 show the unstandardized and standardized coefficients for each predictor. F1 has a positive unstandardized coefficient of 0.048 and is also statistically significant ($t = 5.057$, $p < 0.001$), indicating a strong positive relationship with LogTotal. F2 has a minimal coefficient ($B = 0.008$) and is not statistically significant ($p = 0.895$). F3 shows a coefficient of 0.005, which is also not statistically significant ($p = 0.576$). Correlations show that F1 has the highest zero-order ($r = 0.724$), partial ($r = 0.704$), and part correlation ($r = 0.680$) with LogTotal, further confirming its importance in the model. The VIF values for all predictors are below 1.2, indicating no severe multicollinearity issues. The highest eigenvalue is 1.497, with the lowest being 0.473. The condition index peaks at 1.779, suggesting multicollinearity is not a significant concern.

The regression analysis reveals that the model is statistically significant, with F1 being the most impactful predictor of LogTotal. In contrast, F2 and F3 do not significantly contribute to the model. Collinearity is not a pressing issue, reinforcing the reliability of the model's estimates.

Table 3.
Coefficientsa in 2 type reader prediction.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
2	(Constant)	2.065	0.065		31.852	0.000					
	F1	0.048	0.009	0.717	5.057	0.000	0.724	0.704	0.680	0.899	1.112
	F2	0.008	0.063	0.019	0.133	0.895	0.226	0.026	0.018	0.861	1.161
	F3	0.005	0.009	0.078	0.566	0.576	0.077	0.110	0.076	0.953	1.049

Note: a. Dependent Variable: LogTotal

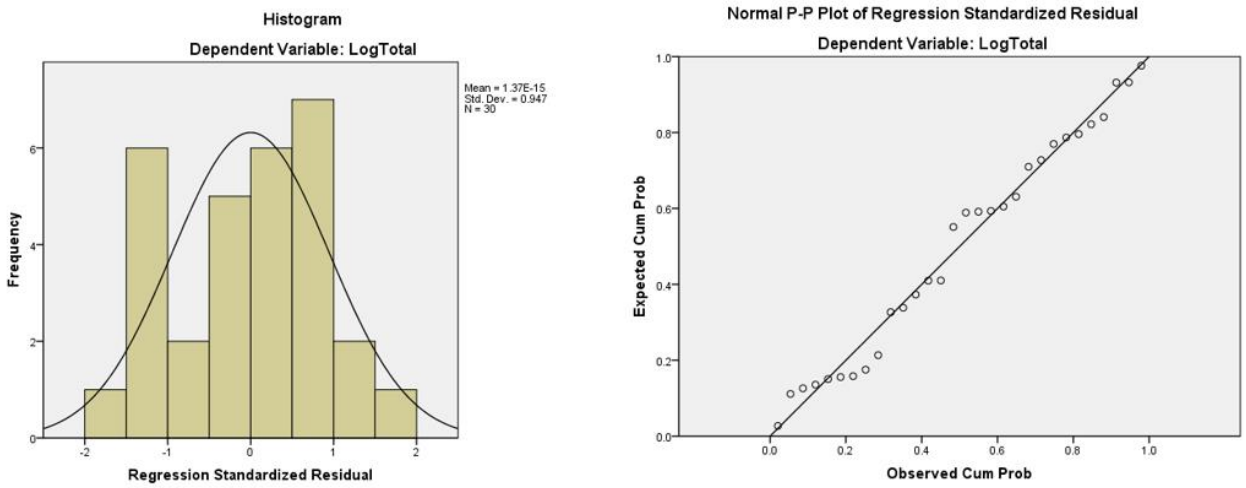

Figure 8.
Histogram and standard P-P plot in type 2 reader prediction.

4.5.3. Type 3 Reader

The regression model demonstrates a moderate correlation between the predictors and the dependent variable (LogTotal), with an R value of 0.480. The R Square value of 0.230 indicates that approximately 23% of the variance in LogTotal can be explained by the three predictors: F1, F2, and F3. The Adjusted R Square is close at 0.229, suggesting the model fits well. The standard error of the estimate is represented by 0.46611, which indicates the average distance that the observed values fall from the regression line. The Durbin-Watson statistic of 1.579 suggests this model likely has no significant autocorrelation issue. Figure 9 displays the histogram and standard P-P plot of the regression standardized residual. The ANOVA table reveals that the regression model is statistically significant, with an F value of 279.352 and a significance level (Sig.) 0.000. This implies that the predictors collectively significantly affect the dependent variable LogTotal. The regression sum of squares is 182.074, while the residual sum is 609.407, contributing to a total sum of squares of 791.481.

The coefficients in Table 4 show the unstandardized and standardized coefficients for each predictor. The constant value is 1.539, with a t-value of 173.911, indicating it is significantly different from zero ($p < 0.001$). The coefficient of F1 is 0.310, with a t-value of 23.431 and $p < 0.001$. This suggests that for each unit increase in F1, LogTotal increases by approximately 0.310 units, holding the other variables constant. The coefficient of F2 is 0.152, with a t-value of 17.656 and $p < 0.001$, implying a positive relationship with LogTotal. A unit increase in F2 is associated with a 0.152 unit increase in LogTotal. The coefficient F3 is 0.064, with a t-value of 5.588 and $p < 0.001$. This signifies a more minor positive effect on LogTotal than the other predictors. The correlations show the strength of the relationships among predictors and the dependent variable, with F1 having the highest correlation with

LogTotal (Beta = 0.390). Collinearity diagnostics indicate that multicollinearity is not a significant concern. The tolerance values for all predictors (F1: 0.992, F2: 0.997, F3: 0.995) are well above the usual threshold of 0.10, and the Variance Inflation Factor (VIF) values are all below 2, indicating that the predictors do not have high multicollinearity.

The regression model provides a robust explanation of the variance in LogTotal, with all predictors demonstrating significant positive relationships with the dependent variable. Each predictor exhibits a positive relationship with LogTotal, confirming their importance in the model. The statistical significance ($p < 0.001$) of all coefficients reinforces the strength of their influence. The regression coefficients indicate varying levels of impact, with F1 demonstrating the most substantial effect, followed by F2 and F3. The absence of multicollinearity issues, as shown by the acceptable tolerance and VIF values, enhances the reliability of the results.

Table 4.
Coefficientsa in 3 type reader prediction.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
3	(Constant)	1.539	0.009		173.911	0.000					
	F1	0.310	0.013	0.390	23.431	0.000	0.368	0.405	0.388	0.992	1.008
	F2	0.152	0.009	0.293	17.656	0.000	0.274	0.316	0.293	0.997	1.003
	F3	0.064	0.011	0.093	5.588	0.000	0.069	0.105	0.093	0.995	1.005

Note: a. Dependent Variable: LogTotal.

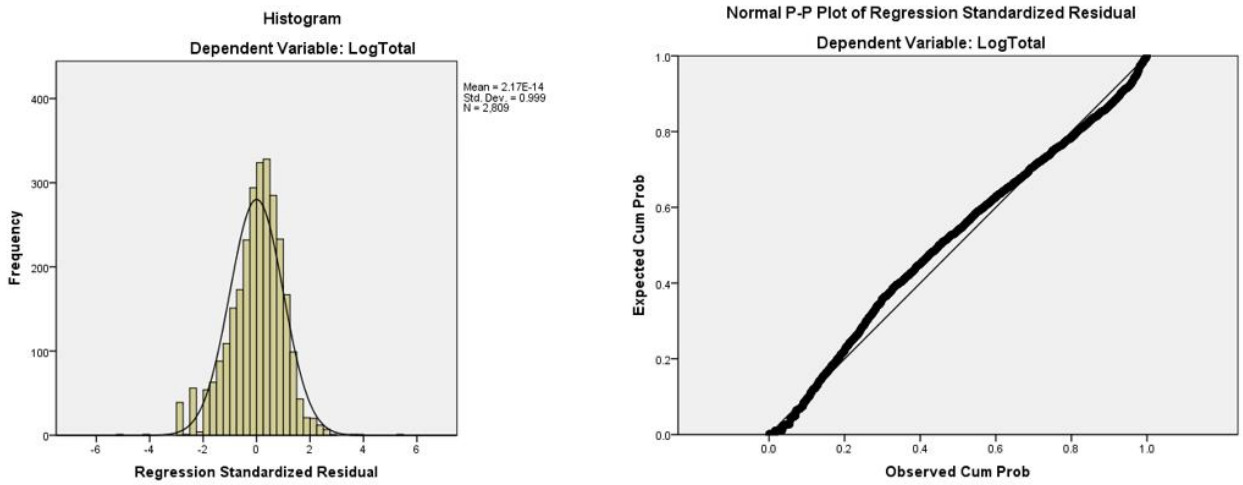

Figure 9.
Histogram and standard P-P plot in type 3 reader prediction.

4.5.4. Unclassified Reader

The correlation coefficient R is 0.447, indicating a moderate positive relationship between the predictors and the dependent variable. The R Square value of 0.199 implies that the three predictors combined can explain approximately 19.9% of the variance in LogTotal. The Adjusted R Square is 0.199, confirming that the model explains a similar proportion of variance. The Durbin-Watson statistic of 1.552 is relatively low, suggesting potential autocorrelation issues. Figure 10 displays the histogram and standard P-P plot of the regression standardized residual. The ANOVA table reveals that the regression sum of squares is 174.744, the Residual Sum of Squares is 701.329, and the total sum of squares is 876.072. The high F-value of 260.622 indicates that the model is statistically significant, and at least one of the predictors is significantly related to the dependent variable ($p < 0.001$).

The coefficients in Table 5 show the unstandardized and standardized coefficients for each predictor. The coefficients (B values) represent the change in LogTotal for a one-unit increase in each predictor. The constant value is 1.549, this is the expected value of LogTotal when all predictors are 0. The coefficient of F1 is 0.188, suggesting that for each unit increase in F1, LogTotal increases by approximately 0.188, with high significance ($p < 0.001$). Similarly, for each unit increase in F2, LogTotal increases by 0.137, also highly significant ($p < 0.001$). This variable has a coefficient of 0.038, indicating a more minor but still significant increase in LogTotal for each unit increase in F3 ($p < 0.001$). The Tolerance and VIF (Variance Inflation Factor) values are 1.000 for all variables, indicating no multicollinearity issues, meaning the predictors do not correlate highly.

The regression analysis suggests that all three predictors (F1, F2, and F3) significantly explain LogTotal variations. The model is statistically significant, but the relatively low R-square indicates that while the predictors provide explanatory power, a significant portion of the variance remains unexplained.

Table 5.
Coefficients[a] in unclassified reader prediction.
Note: a. Dependent Variable: LogTotal.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
4	(Constant)	1.549	0.008		183.715	0.000			1.549	0.008	
	F1	0.188	0.008	0.356	22.290	0.000	0.356	0.370	0.188	0.008	0.356
	F2	0.137	0.008	0.260	16.276	0.000	0.260	0.279	0.137	0.008	0.260
	F3	0.038	0.008	0.072	4.483	0.000	0.072	0.080	0.038	0.008	0.072

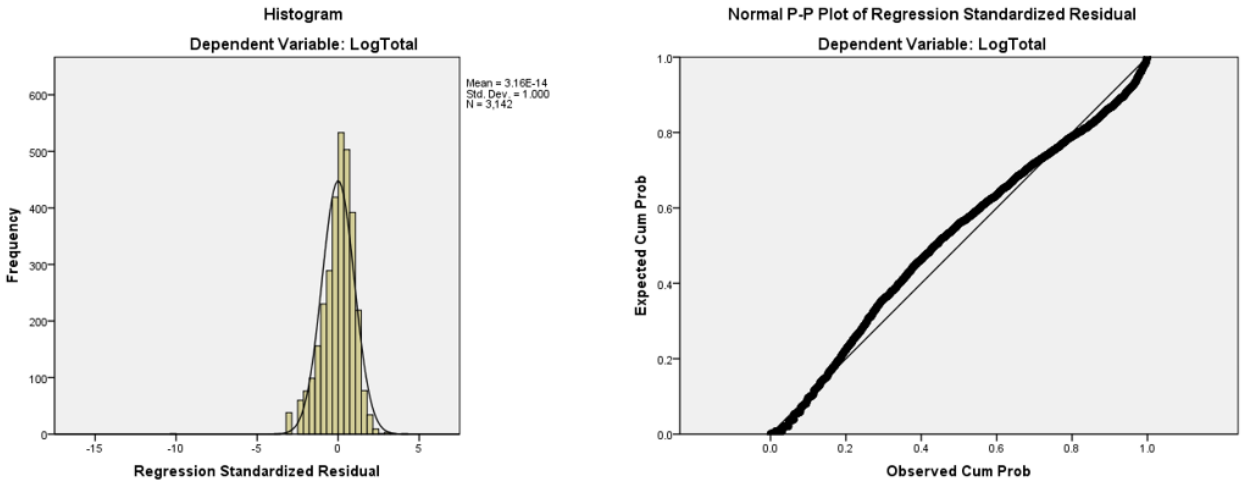

Figure 10.
Histogram and standard P-P plot in unclassified reader prediction.

5. Discussion

5.1. Popular Subject for Statistics Books

Analyzing popular statistics books in the library of Nanjing Normal University based on subject headings and word frequency can provide valuable insights into readers' academic preferences and priorities. This dual-pronged approach offers a comprehensive understanding of both the organizational structure of the literature available and the content that engages users, facilitating better resource allocation, collection development, and tailored educational initiatives in the field of statistics.

The popularity of statistical books in normal universities can be attributed to several interconnected factors and emerging trends. In an era where data-driven decision-making is paramount across numerous fields, there is a growing emphasis on equipping students with statistical literacy. As industries increasingly rely on data analysis for insights and strategies, universities prioritize teaching statistical concepts to prepare graduates for the workforce. The findings indicate strong statistical applications in diverse fields, such as medical and social statistics. This cross-disciplinary approach enhances the relevance of statistics, making it an essential component of curricula in areas ranging from public health to social sciences. The popularity of statistical software (like SPSS and Excel) has revolutionized how statistics is taught and applied. These tools facilitate practical learning and enable students to handle large datasets efficiently. Hence, textbooks often focus on real-world applications, integrating software training with theoretical concepts. With the rise of empirical research in academia, there is an increasing emphasis on teaching robust statistical methodologies. This is evident in the prominence of multivariate analysis, regression analysis, and other advanced techniques in university coursework. As students engage in research projects, access to comprehensive resources on these methods is crucial. The commitment to providing resources for educators and graduate students indicates a strategic focus on enhancing teaching methods and learning experiences. Textbooks that align with pedagogical approaches and support a variety of educational levels contribute to a thriving academic environment. Contemporary topics like machine learning and Bayesian statistics reflect universities' efforts to stay current with the latest field advancements. As new statistical methods gain traction, academic institutions adapt curricula to encompass these trends, ensuring students are well-equipped for current and future challenges in data science. Many universities collaborate with various industries and organizations, leading to curriculum changes that emphasize the practical implications of statistics. This partnership helps align educational resources with the job market's needs, driving demand for statistical textbooks that cover applied concepts. Therefore, the popularity of statistical books in normal universities is a multifaceted phenomenon driven by the rising importance of data literacy, interdisciplinary integration of statistical methods, technological advancements, and a forward-thinking approach to curriculum development. These trends indicate a robust commitment to preparing students for an increasingly data-oriented world.

5.2. Book-Reader Correspondence Relationships

Examining the Book-Reader correspondence elucidates critical relationships between reader preferences and book classifications. The analysis unveils distinct engagement trajectories among the identified reader groups. Reader Group 3 emerges as the most actively engaged cohort, demonstrating substantial interactions across various book categories, particularly with Book Group 1. This observation signifies a pronounced affinity for the types of literature encompassed within this classification. The aggregate scores reveal significant engagement levels, as Reader Group 1 attaining 2393 from Book Group 1 indicates considerable interest. In contrast, Reader Group 3 garners the highest engagement score 1962 from Book Group 3. This juxtaposition suggests that Reader Group 3 exhibits considerable engagement while Reader Group 1 maintains a noteworthy interest, particularly within specific book classifications. The correspondence analysis elucidates that Dimension 1 accounts for a substantial portion of the variance (0.988), indicating its capacity to capture most relational dynamics among the reader groups and book classifications. The elevated inertia (0.951) associated with Reader Group 1 in Dimension 1 signifies its critical role in delineating the observed relationships, suggesting that the preferences of this group exert a considerable influence on the variance in book engagement. Conversely, Reader Group 2 exhibits a relatively minimal impact on Dimension 1 (0.061), indicating a more specialized interest that may not align closely with the broader preferences of the dataset. This observation suggests that the reading habits of this group are more niche in nature, presenting challenges for widespread engagement across different book classifications. These insights have significant implications for engagement strategies among stakeholders. Targeted promotions for Book Groups 1 and 3 directed at Reader Group 1 could potentially enhance user engagement by

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

aligning offerings with their demonstrated preferences. Similarly, a nuanced understanding of the unique reading interests of Reader Group 2 could facilitate the development of more focused and effective outreach initiatives. The findings from this correspondence analysis underscore the necessity of recognizing the heterogeneity of reader preferences. Tailoring initiatives based on these differentiated reading habits is imperative for strengthening engagement with literary resources, thereby ensuring that offerings resonate meaningfully with the specific interests of each reader group.

5.3. Reader Clusters and Factors

The analysis of reader clusters and influencing factors reveals distinct patterns in borrowing behavior and engagement levels within library user groups, as categorized by key parameters. The analysis reveals important insights about borrower engagement and preferences, which can inform library services and collection development.

Borrower clusters: Cluster 1 comprises 303 borrowers with lower engagement and borrowing frequency. This cluster most likely represents casual readers who engage infrequently with library resources. It indicates a significant, albeit smaller, population segment that may benefit from targeted outreach initiatives to promote library resources and encourage increased usage. Cluster 2 with only 30 borrowers, reflects a high level of engagement with library materials. Despite its relatively small size, this cluster could represent a dedicated group of readers with specific interests or niche preferences. Targeted recommendations and specialized collections could enhance the experience of this group, fostering further engagement and satisfaction. In contrast, Cluster 3 represents the largest demographic with 2809 borrowers, demonstrating moderate engagement with library resources. This suggests a robust base of users who regularly utilize library services, indicating an opportunity for programs and services that cater to the general reader.

Factor analysis: F1 encompasses social and fundamental natural sciences and significantly predicts borrowing behavior (LogTotal) across all reader types. Its prominence suggests a strong interest in materials that intertwine social and scientific inquiries, aligning with contemporary academic trends and interdisciplinary research. F2 focuses on socio-political and historical contexts and exhibits varying significance among reader types. For type 1 and type 2 readers, F2 does not contribute significantly to borrowing behavior, indicating a lesser prioritization of socio-political or historical content. However, for type 3 readers, F2 demonstrates a moderate influence, suggesting that this demographic possesses some level of engagement with socio-political themes, albeit not as a primary driver. F3 pertains to environmental and agricultural disciplines and appears to be the least influential across all reader types. While environmentally conscious readers may be interested, the findings imply that it does not significantly impact broader borrowing behavior within the library population.

Reader types: Type 1 Readers are influenced primarily by F1 and F3, with F2 playing no significant role in their borrowing decisions. This suggests a preference for interdisciplinary works that connect social and natural sciences. Type 2 Readers strongly rely on F1 as the predominant factor influencing their borrowing choices. The lack of notable contributions from F2 and F3 indicates that this group favors straightforward scientific content over socio-political discussions or agricultural themes. Type 3 readers exhibit F1 as the strongest predictor of borrowing behavior, with F2 having a positive contribution but with a lower effect than F1. F3 exerts the least influence. This indicates that type 3 readers may possess a more eclectic range of interests, although they remain primarily driven by themes within social sciences and natural sciences.

Analyzing borrower clusters and factors influencing reader engagement provides valuable information for library administrators and policymakers. Targeted strategies aimed at enhancing the experience of Cluster 1 and catering to the niche interests of Cluster 2 could yield positive outcomes in terms of increased borrowing frequency and user satisfaction. Moreover, leveraging the interdisciplinary nature of F1 across various reader types may foster more profound engagement with library collections, ultimately enhancing the overall user experience.

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

5.4. Comparison of Prediction Results Between Integrated and Non-Integrated Algorithms

Model 2 is the most effective at predicting LogTotal, showing the strongest correlation and the highest explanatory power, followed by Model 1. Models 3 and 4 demonstrate weaker performance overall, especially with Model 4 indicating potential issues with autocorrelation. Each model's metrics provide valuable insights into how well they fit the data and their predictive capabilities. The clustering regression integration algorithm used in Models 1, 2, and 3 provides distinct advantages over the algorithm employed in Model 4. There are some of the key benefits of using clustering regression:

Enhanced Predictive Accuracy: Clustering regression integration can improve predictive accuracy by identifying and utilizing the underlying patterns in the data. By grouping similar data points, the model can tailor its predictions based on distinct segments or clusters, leading to more precise results.

Handling Heterogeneity: The clustering approach allows the model to handle heterogeneous data better. Different clusters may have different relationships between predictors and the dependent variable. This adaptability improves performance, especially in complex datasets where relationships may not be uniform across all observations.

Robustness to Outliers: Clustering methods can be more robust to outliers, focusing on group characteristics rather than individual points. By entering the model around clusters, the impact of outliers on the overall regression analysis can be minimized, resulting in a more reliable model.

Improved Interpretability: The clustering regression approach allows for more interpretable results. Each cluster can be analyzed separately, making it easier to understand how different data segments contribute to the overall relationship with the dependent variable. This is especially useful for informing decision-making and operational strategies.

Increased Flexibility: Models incorporating clustering can adapt more flexibly to various patterns in the data. This can include nonlinear relationships that may be overlooked in traditional regression approaches, leading to more nuanced insights about the predictors' effects.

Feature Engineering: Clustering can aid in feature engineering by allowing the inclusion of cluster-related variables. This can enhance the model's performance by providing additional relevant information that might not be captured through straightforward regression methods alone.

Segmentation Insights: Clustering regression can help identify segments within the data that behave similarly, enabling targeted strategies based on those segments. This can be particularly advantageous in library applications, where understanding reader segments is crucial.

The clustering regression integration algorithm improves accuracy and interpretability by focusing on the relationships within distinct data segments. This contrasts with a more straightforward integration approach in Model 4, which may not leverage these advantages, leading to lower predictive power and interpretability. The clustering approach is especially beneficial for complex, heterogeneous datasets where relationships are not straightforward.

6. Conclusion

This research contributes valuable insights that can guide future curriculum development and outreach strategies in statistical education. the analysis of the popularity of statistical books at Nanjing Normal University reveals a complex interplay of factors that shape reader preferences and engagement. The findings underscore the significance of data literacy and the interdisciplinary application of statistical methods, reflecting the evolving educational landscape. Reader Group 3 stands out for its active engagement, particularly with Book Group 1, suggesting that strategic promotions targeting this group could enhance overall user interaction. Moreover, the distinct reading habits of Reader Group 1 and Reader Group 2 highlight the necessity for tailored outreach initiatives. Type 1 Readers are inclined toward interdisciplinary works, while Type 2 Readers prefer direct scientific content, indicating the diverse nature of reader interests. Type 3 Readers, with their eclectic preferences, further emphasize the need for a broad yet focused approach to curating and promoting statistical literature. Lastly, applying a clustering regression integration algorithm demonstrates its effectiveness in improving accuracy and interpretability by better understanding the dynamics within varied data segments. This innovative

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

approach contrasts favorably with traditional methods, affirming the need for advanced analytical techniques to capture the intricacies of reader behavior.

While informative, the study's conclusions regarding book borrowing patterns among different majors at Nanjing Normal University Library are subject to several limitations. The data is derived from the Nanjing Normal University Library, which may not represent borrowing trends across various universities or libraries with different curricular emphases. For instance, institutions with strong interdisciplinary programs may show different borrowing patterns, indicating that students from multiple disciplines often use statistics resources. The study covers a specific period from 2014 to 2023. The book circulation data may not fully accommodate all access forms to statistics materials, including digital resources, open educational resources, or collaborative projects that might involve statistics. This oversight may lead to underrepresenting the engagement levels of interdisciplinary students who draw knowledge from multiple domains. Changes in interdisciplinary program offerings or the integration of statistics into various fields of study might influence borrowing patterns significantly over time. For example, emerging interdisciplinary fields such as data science may drive increased interest in statistics across non-traditional disciplines. Classifying students into specific major categories does not account for interdisciplinary studies or major shifts. Students engaging in combined or dual degree programs may borrow more statistics books for their varied coursework, which may not accurately reflect. Various external factors, such as advancing technology and evolving job markets, may create new interdisciplinary connections that influence students' interest in statistics. For example, as fields like psychology increasingly utilize data analytics, students in that major might borrow more statistics materials, suggesting that a cross-disciplinary approach is becoming more common. Interdisciplinary students might borrow less frequently due to their diverse resource needs but may engage with statistics in various ways not captured by circulation data alone.

Cluster size is critical in influencing the results of linear regression analyses. Larger clusters can provide more stable estimates, better predictive power, and excellent overall reliability. However, the nature and relevance of the variables chosen also significantly impact model performance, highlighting the importance of not just size but also the appropriate selection and treatment of predictors. The number of observations can affect the model's performance and interpretability. Larger clusters, like Cluster 3 with 2809 borrowers, often yield more reliable and stable regression estimates. This is because a larger sample size generally provides greater statistical power, allowing for more precise estimates of parameters and reducing the likelihood of Type 1 and Type 2 errors. In contrast, smaller clusters like Clusters 1 and 2 (303 and 30 borrowers, respectively) may exhibit more variability in results, and any conclusions drawn from them should be made cautiously. The regression results for Cluster 2 indicate a robust R value of 0.728 and an R Square of 0.530, suggesting that the predictors can explain almost half of the variability in the dependent variable (LogTotal). This strong relationship could be partly because of the sufficient size of the cluster, allowing for a more accurate reflection of underlying trends. In contrast, Cluster 1, despite being more significant than Cluster 2, has lower R and R-Square values (0.590 and 0.348), highlighting that not all larger clusters necessarily provide more substantial predictive power. The Adjusted R-Square, which accounts for the number of predictors in the model, remains significant across clusters, but smaller clusters may have less stable and potentially skewed adjustments. For instance, Cluster 3 has an R Square of 0.230, indicating that the model explains less of the variance despite its size. This suggests that while the sample size is large, the predictors may not be as relevant for this group, pointing to the need for more tailored models or additional predictors. The standard error of the estimate offers insights into the accuracy of predictions. Smaller clusters may have larger standard errors due to the inherent variability and fewer data points, making predictions less reliable. In contrast, larger clusters can provide minor standard errors, reflecting more consistent relationships between variables. The Durbin-Watson Statistic is important for assessing autocorrelation in residuals. While values in all clusters suggest minimal autocorrelation, the larger sample in Cluster 3 gives a bounce to model stability. Nonetheless, moderate autocorrelation in Clusters 2 and 3 suggests that more complexities in those data structures may need further examination. Larger clusters tend to

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-26, 2025
DOI: 10.55214/26410230.v7i2.6358

yield models with greater predictive validity, as evidenced by Cluster 2's strong statistical significance (p-value of 0.000). Smaller clusters may not achieve similar significance levels, thus impacting the ability to make generalizations from those models.

By considering these limitations, mainly through an interdisciplinary lens, we can better contextualize the results and understand the broader factors that may influence book borrowing behavior in various academic disciplines. This approach emphasizes the interconnectedness of knowledge and the importance of cross-disciplinary engagement in education.

Transparency:

The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] J. L. Díaz Palencia, "Applying the anthropological theory of didactics to enhance a multicultural approach to statistics education for engineering students," *Journal for Multicultural Education*, vol. 18, no. 4, pp. 435-446, 2024. https://doi.org/10.1108/JME-04-2024-0043

[2] T. C. Nokeri, "Introduction to econometrics. In: Econometrics and Data Science." Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-7434-7_1, 2022.

[3] J. B. Ärlebäck and T. Kawakami, *The relationships between statistics, statistical modelling and mathematical modelling* (Advancing and consolidating mathematical modelling: Research from ICME-14). Springer, 2023.

[4] M. Kadosh, M. Hen, and J. R. Ferrari, "Reducing statistics anxiety and academic procrastination among Israeli students: A pilot program," *Teaching Statistics*, vol. 45, no. 3, pp. 167-175, 2023. https://doi.org/10.1111/test.12356

[5] R. J. Rossi, *Applied biostatistics for the health sciences*. John Wiley & Sons. https://doi.org/10.1002/9781119722717.ch5, 2022.

[6] H. T. Nguyen, "A prelude to statistics arising from optimal transport theory," *Asian Journal of Economics and Banking*, vol. 7, no. 2, pp. 166-179, 2023. https://doi.org/10.1108/AJEB-05-2023-0038

[7] J. Carbonara and E. Fokoue, "Augmenting intelligence: The convergence of ML/LLMs and statistics," *Stat*, vol. 14, no. 1, p. e70043, 2025. https://doi.org/10.1002/sta4.70043

[8] C. O. Obazuaye, "Advanced statistics training for HRD doctoral students: a call for empirical study!," *European Journal of Training and Development*, 2025. https://doi.org/10.1108/EJTD-08-2024-0112

[9] C. Wang, "International vital statistics: A resource guide," *Collection Building*, vol. 22, no. 2, pp. 53-59, 2003. https://doi.org/10.1108/01604950310469976

[10] A. Brisbin and E. Maranhao do Nascimento, "Reading versus doing: Methods of teaching problem-solving in introductory statistics," *Journal of Statistics Education*, vol. 27, no. 3, pp. 154-170, 2019. https://doi.org/10.1080/10691898.2019.1637801

[11] M. A. Sole and S. L. Weinberg, "What's Brewing? A Statistics Education Discovery Project," *Journal of Statistics Education*, vol. 25, no. 3, pp. 137-144, 2017. https://doi.org/10.1080/10691898.2017.1395302

[12] T. A. Forest, M. L. Schlichting, K. D. Duncan, and A. S. Finn, "Changes in statistical learning across development," *Nature Reviews Psychology*, vol. 2, no. 4, pp. 205-219, 2023.

[13] R. Heijungs, *Statistics 2: Inferential. In Probability, Statistics and Life Cycle Assessment*. Cham. : Springer. https://doi.org/10.1007/978-3-031-49317-1_5, 2024.

[14] L. A. Hildreth, M. Miley, E. Strickland, and J. Swisher, "Writing workshops to foster written communication skills in statistics graduate students," *Journal of Statistics and Data Science Education*, vol. 31, no. 2, pp. 201-210, 2023. https://doi.org/10.1080/26939169.2022.2138800

[15] R. Boyask, J. Jackson, J. Milne, C. Harrington, and R. May, "We enjoy doing reading together: Finding potential in affective encounters with people and things for sustaining volitional reading," *Language and Education*, vol. 38, no. 4, pp. 578-595, 2024. https://doi.org/10.1080/09500782.2024.2337658

[16] A. Cujba and M. Pifarré, "Enhancing students' attitudes towards statistics through innovative technology-enhanced, collaborative, and data-driven project-based learning," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1-14, 2024. https://doi.org/10.1057/s41599-024-03469-5

[17] W. Wang, *Statistical framework. In Principles of Machine Learning: The Three Perspectives*. Singapore: Springer Nature Singapore, 2024.

[18] J. Hedderich and L. Sachs, *Descriptive statistics* (Applied Statistics). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-70074-7_3, 2024.

[19] C. Starbuck, *Descriptive Statistics. In: The Fundamentals of People Analytics*. Cham: Springer. https://doi.org/10.1007/978-3-031-28674-2_7, 2023.

[20] D. Selvamuthu and D. Das, *Descriptive statistics* (Introduction to Probability, Statistical Methods, Design of Experiments and Statistical Quality Control). Singapore: Springer, 2024.

[21] T. Cleff, *Statistics and empirical research* (Applied Statistics and Multivariate Data Analysis for Business and Economics: A Modern Approach Using R, SPSS, Stata, and Excel). Cham: Springer, 2025, pp. 1-15.

[22] C. A. Saliya, *Econometrics. In: Doing Social Research and Publishing Results*. Singapore: Springer. https://doi.org/10.1007/978-981-19-3780-4_14, 2022.

[23] F. J. Bismans and O. Damette, *Dynamics in econometrics. In Dynamic Econometrics*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-72910-2_2, 2025.

[24] G. Arbia, *Further topics in spatial econometrics. In: A Primer for Spatial Econometrics. Palgrave Texts in Econometrics*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-57182-4_4, 2024.

[25] B. H. Baltagi, *What is econometrics? In Solutions Manual for Econometrics: Classroom Companion: Economics*. Cham: Springer. https://doi.org/10.1007/978-3-030-80158-8_1, 2022.

[26] E. N. Barron and J. G. Del Greco, *Confidence and prediction intervals. In: Probability and Statistics for STEM. Synthesis Lectures on Mathematics & Statistics*. Cham: Springer. https://doi.org/10.1007/978-3-031-38985-6_4, 2024.

[27] S. K. Sahu, *Introduction to basic statistics* (Introduction to Probability, Statistics & R: Foundations for Data-Based Sciences). Cham: Springer, 2024.

[28] S. Prasad, *Vital statistics. In Advanced Statistical Methods*. Singapore: Springer. https://doi.org/10.1007/978-981-99-7257-9_4, 2024.

[29] R. Sen and S. Das, *Descriptive statistics* (Computational Finance with R). Singapore: Springer, 2023.

[30] K. Ooka and M. Arai, "Accurate prediction of protein folding mechanisms by simple structure-based statistical mechanical models," *Nature Communications*, vol. 14, no. 1, p. 6338, 2023. https://doi.org/10.1038/s41467-023-41664-1

[31] A. Sharaff, A. Sonkusare, A. Pal, and S. Pahadi, "Statistical modeling of temperature prediction using residual network," in *International Conference on Computing, Communication and Learning*, 2024: Springer, pp. 360-372.

[32] I. Vasilev, "Developing a library of tree-based models for survival analysis," *Moscow University Computational Mathematics and Cybernetics*, vol. 48, no. 3, pp. 190-202, 2024. https://doi.org/10.3103/S0278641924700134

[33] D. Jangir, L. Hota, B. P. Nayak, and A. Kumar, "Football match result prediction using twitter statistical/historical data," in *International Conference on Deep Learning, Artificial Intelligence and Robotics*, 2023: Springer, pp. 220-236.

[34] N. Wu, "Library reader behavior based on apriori association algorithm," in *Innovative Computing: Proceedings of the 4th International Conference on Innovative Computing (IC 2021)*, 2022: Springer, pp. 1095-1102.

[35] Y. Chen, L. Xing, and J. Liu, "Data analysis on library entry behavior of university library," in *Innovative Computing: Proceedings of the 4th International Conference on Innovative Computing (IC 2021)*, 2022: Springer, pp. 165-172.

[36] Y. Quan Liu, "A comparison of public library use of statistics–a cross-country survey report," *New Library World*, vol. 102, no. 4/5, pp. 138-146, 2001. https://doi.org/10.1108/03074800110390545

[37] D. Kallivokas, "Learning statistics using specialized software applications," in *The International Conference on Strategic Innovative Marketing and Tourism*, 2023: Springer Nature Switzerland Cham, pp. 749-756.

[38] A. Khandare, N. Agarwal, A. Bodhankar, A. Kulkarni, and I. Mane, *Analysis of python libraries for artificial intelligence* (Intelligent computing and networking: Proceedings of ic-icn 2022). Singapore: Springer, 2023.

[39] P. Meesad and A. Mingkhwan, *Algorithmic innovations: Pioneering the future of library services* (Libraries in Transformation: Navigating to AI-Powered Libraries). Cham: Springer, 2024, pp. 57-97.

[40] L. Opstad, "Is performance in mathematics and statistics related to success in business education?," *Journal of Applied Research in Higher Education*, vol. 16, no. 5, pp. 1925-1936, 2024. https://doi.org/10.1108/JARHE-08-2023-0361

[41] D. Giuliani, M. M. Dickson, and G. Espa, "Teaching statistics in the context of social foresight. An applied approach based on the use of an open-source software," *On the Horizon*, vol. 23, no. 2, pp. 140-148, 2015. https://doi.org/10.1108/OTH-02-2015-0010

[42] D. A. Yousef, "Determinants of the academic performance of undergraduate students in statistics bachelor's degree program: A study in an Arabic setting," *Quality Assurance in Education*, vol. 27, no. 1, pp. 13-23, 2019. https://doi.org/10.1108/QAE-12-2016-0087

[43] H. Dodeen and S. Alharballeh, "Predicting statistic anxiety by attitude toward statistics, statistics self-efficacy, achievement in statistics and academic procrastination among students of social sciences colleges," *Journal of Applied Research in Higher Education*, 2024. https://doi.org/10.1108/JARHE-01-2024-0021