The Application of Non-parametric Statistics in Consumer Preference Analysis and Precision Marketing within the Context of the New Retail Business Model

Haoyang Cheng

University of California, Irvine, The United States of America

haoyac9@uci.edu

Keywords: New retailing; Consumer preference analysis; Nonparametric statistics; Kernel density estimation

Abstract: As Internet technology develops and big data becomes prevalent, new retail business models gradually transform consumer habits and business operation modes. New retailing emphasizes the integration of online and offline, relying on data analysis to achieve precision marketing and consumer preference analysis. This paper explores the application of nonparametric statistical methods in new retailing, analyzing consumer behavior and optimizing marketing strategies. Nonparametric statistics overcomes the limitations of traditional parameter statistics on distribution assumptions. Furthermore, it applies to situations where data characteristics are unknown or do not conform to a specific distribution. This study uses nonparametric methods, including kernel density estimation, rank sum test, and Kruskal-Wallis H test, to analyze the transaction data from an e-commerce platform. The results indicate significant regional and temporal characteristics in consumer preferences. Additionally, this study employs the Bootstrap method to assess the impact of marketing activities and proposes corresponding precision marketing strategies. The research findings demonstrate that the consumer preference analysis realized by nonparametric statistical methods can effectively guide retailers to formulate precision marketing strategies to improve user experience and sales performance.

1. The Basis of Non-parametric Statistical Methods

1.1 An Overview of Non-parametric Statistical Methods

Nonparametric statistical methods are statistical analysis methods that do not depend on the specific parameter settings of the data distribution. When analyzing statistical data, nonparametric methods do not assume any specific distribution, such as normal distribution or binomial distribution. This flexibility makes nonparametric methods highly applicable in practical situations, especially when dealing with small sample sizes or when the data deviates from common distribution patterns. Nonparametric methods can provide effective insights into data structures and trends, especially for consumer behavior analysis in modern business areas such as new retailing, because these data are often highly nonlinear, influenced by complex factors, and their distribution is unknown [1].

Nonparametric methods generally have fewer prerequisites compared to parametric methods. They do not require an accurate prediction of the data's distribution characteristics. Instead, these methods focus on analyzing and inferring insights directly from the arrangement and distribution of the data. For instance, when assessing the sentiment trends in consumer reviews, it is challenging to determine the precise type of sentiment distribution beforehand. In such cases, nonparametric methods demonstrate their advantages in the analysis.

1.2 Introduction of Common Non-parametric Statistical Methods

Kernel Density Estimation (KDE): KDE is a method for estimating the probability density function. In new retailing, KDE can assist retailers in ascertaining the probability distribution of consumer purchase behavior. For instance, it can be used to determine the intensity of consumers' preferences for a certain type of product. KDE is capable of not only dealing with single-dimensional data, but also extending to multi-dimensional data to reveal the joint distribution of multiple variables [2].

  DOI: 10.25236/gemmsd.2025.024

Rank Sum Test (Wilcoxon Test): The Rank Sum Test is a nonparametric method used to determine whether two independent samples come from the same distribution [3]. In new retailing, the rank sum test can compare overall preferences between two consumer groups, such as those from different regions, age groups, or genders, regarding a product.

Kruskal-Wallis H Test: The Kruskal-Wallis H Test is a nonparametric analysis of variance used to test whether three or more independent samples are from the same distribution. In the new retailing business, this method can be used to compare whether different user groups (such as different levels of members) show the same distribution of product satisfaction.

Mann-Whitney U Test: The Mann-Whitney U test is a type of rank sum test designed for two independent samples. It is used to determine whether the two data sets show a significant difference in their medians. For instance, this test can be applied to analyze whether different marketing strategies lead to significantly different sales results [4].

2. Data Collection of Consumer Preference with New Retail Business Model

2.1 The Characteristics of Consumer Data from the Perspective of New Retailing

In the new retail model, consumer data collection is marked by a significant level of diversity and complexity. These data come from customers' online shopping habits, offline shopping behaviors, social media interactions, customer service conversations, and various life application scenarios [5]. Consumer preference data covers consumption history and reflects consumer behavior habits, lifestyles, and consumption psychology. These data usually have the following characteristics:

Massive: With the popularity of smart devices, consumers' digital footprints have multiplied, providing massive data for analysis.

High-dimensional: The data explores consumers' behaviors across multiple dimensions, such as time, place, frequency, and costs.

Unstructured: Many data, such as comments on social media and product reviews, are unstructured in text form.

Real-time: In the new retailing model, data is generated almost in real time, including real-time purchases from e-commerce platforms and immediate feedback on social networks.

2.2 Data Collection Channels and Methods

In the new retailing, effective data collection channels mainly include:

E-commerce platform: Information about online shopping behavior is crucial for understanding consumer preferences, encompassing amounts of details like browsing history, purchase history, and items in the shopping cart.

Mobile applications: The consumption data collected through mobile payment, location services, and other applications can help us analyze the relationship between consumer preferences and mobile behavior.

Social media: Consumers' discussion of brands and evaluation of products reflect their direct feelings and preferences [6].

IoT devices: Data generated by IoT devices such as smart homes and wearable devices can reflect consumers' living habits and potential needs.

Data collection approaches include but are not limited to:

API acquisition: API acquisition is a good way to use the application programming interface provided by the third party to collect users' public data.

Web crawler: For online platforms such as social media and forums, it is suggested that data be collected using automatic crawler tools.

POS system: Sales data of offline retail outlets are collected through the POS system and integrated with online data.

User feedback: Direct user feedback, such as a questionnaire survey, telephone interview, or face-to-face communication, is a good way to get feedback.

2.3 Data Preprocessing

Data preprocessing is crucial in data analysis, and its purpose is to clean, transform, and standardize the original data so that the subsequent statistical analysis can be carried out effectively [7]. It usually involves the following steps:

Data cleaning: It includes removing duplicate records, filling missing values, and eliminating abnormal values. These operations are conducive to improving the quality of the data set.

Format standardization: Given the heterogeneity of the data collected from various sources, standardizing the formats is imperative to ensure seamless analysis.

Processing for noise data: In consumer data, noise signals will inevitably appear, and the interference information will be minimized by smoothing and filtering.

Text processing: Unstructured text data must be transformed into structured data for statistical analysis using text mining techniques, such as word frequency statistics, sentiment analysis, and topic modeling.

Data normalization: To normalize data of different scales to the same standard to avoid errors caused by different scales.

Data integration: When integrating data from different data sources and channels, it is necessary to ensure the consistency and integrity of the data.

After the pretreatment mentioned above, the consumer preference data in new retailing will become more accurate, consistent, and of analytical value. Nonparametric statistical methods will help uncover more profound consumer preferences and enhance precision marketing effectiveness.

3. Consumer Preference Analysis Based on Nonparametric Statistical Methods

3.1 Analysis of the Relationship between Consumers' Basic Characteristics and Preferences

For the new retail business model, using nonparametric statistical methods to analyze the relationship between consumers' basic characteristics and preferences is key to gaining accurate market insight. By collecting basic information such as consumers' age, gender, occupation, and income level, we can use the Spearman rank correlation coefficient to evaluate the correlation between these basic characteristics and consumers' purchasing preferences. Using kernel density estimation, we can analyze the distribution of consumer preferences for a specific commodity category or brand and assess whether significant differences exist among various consumer groups through a rank sum test. By using these methods, we do not need to assume the data's distribution, allowing for a more accurate reflection of the actual relationship between diverse features and preferences.

3.2 Analysis of Consumer Purchase Behavior Preferences

The analysis of purchasing behavior preference focuses on when, where, how, and why consumers buy products. Non-parametric statistics can effectively analyze purchase behavior data and uncover potential consumer preferences. For instance, the Mann-Whitney U test can compare consumer buying behavior differences between weekdays and weekends. The Kruskal-Wallis H test is used to analyze the differences in purchase frequency and costs spent between consumers with different consumption levels (such as VIP level). These analyses assist retailers in understanding consumer buying patterns and habits, allowing them to optimize product displays, inventory management, and promotional activities.

3.3 An Analysis of Consumers' Preference for Marketing Activities

Marketing activity is an important means to attract consumers' attention and stimulate purchases in the new retailing model. By utilizing nonparametric statistics to analyze consumer responses and preferences regarding various marketing activities, retailers can determine which strategies attract customers and enhance the conversion rate. We can draw multiple sample datasets from successful historical marketing activities for resampling by applying the nonparametric Bootstrap method. This allows us to estimate the confidence interval for the impact of marketing activities on consumer behavior. It helps to evaluate the potential effect of the newly proposed marketing scheme and adjust the strategy to target specific consumer groups.

3.4 Analysis of Dynamic Change of Consumer Preferences

Consumer preferences are dynamic and will adjust over time, as well as market trends. Using nonparametric statistical methods, such as the coefficient of variation and Spearman rank correlation coefficient, we can analyze the fluctuation of consumer preference with time and observe the periodic change of purchasing intention intensity. By analyzing the changes in consumer buying behavior before and after holidays, we can better understand the specific impact of holiday marketing on consumer preferences and purchasing decisions. Similarly, by continuously monitoring consumer activities and feedback on social media and employing time series analysis, we can track the shifting trends in consumer preferences for particular products or services. This information can serve as a foundation for making timely adjustments to marketing strategies. For example, by comparing the dynamic changes in consumers' buying behavior before and after holidays, we can explore the specific influence of holiday marketing on consumers' preferences and buying behavior. Similarly, by continuously monitoring consumers' activities and feedback on social media, combined with time series analysis, we can track the trend changes in consumers' preferences for specific products or services and provide a basis for timely adjustment of market strategies.

To sum up, nonparametric statistical methods analyze consumers' basic characteristics, purchasing behavior, response to marketing activities, and dynamic changes in preferences. It provides detailed market insights and data-driven decision support for merchants in the new retail sector. The flexibility and effectiveness of these methods serve as valuable tools for understanding and addressing consumers' diverse and rapidly changing needs. Through careful analysis of consumers' preferences, new retailing merchants will more accurately locate the target market, formulate effective marketing strategies, and enhance the customer experience to remain competitive in the complex and changeable market environment.

4. Empirical Study

4.1 Research Objects and Data Sources

4.1.1 Research Objects

The empirical research of this paper focuses on a large supermarket chain operating under a new retail business model. The supermarket integrates online and offline sales channels, provides a variety of goods and services, and has the typical characteristics of new retailing. The aim of this study is to examine supermarket consumer preferences to provide a foundation for precision marketing.

4.1.2 Data Sources

Online data: Gather consumers' browsing history, purchase records, and collection details from supermarket official websites and mobile applications. These data cover over 50,000 transaction records from last year, including basic consumer information (age, gender, region), category, brand, and price.

Offline data: Collect offline consumer data, such as consumption time, amount spent, and product mix, from supermarket member systems and cashier records. In addition, questionnaire survey points were set up in the supermarket. Consumers were randomly selected for the questionnaire survey, and 2000 valid questionnaires were collected. The survey's content involves consumers' shopping habits, preferences, and feedback on marketing activities.

4.2 The Application Process of Non-parametric Statistical Methods

4.2.1 Data Preprocessing

We clean and integrate the collected online and offline data to remove duplicates, errors, and missing data. The age of consumers is divided into different age groups (such as 18-25 years old, 26-35 years old, 36-45 years old, etc.), and the commodity categories are classified.

4.2.2 An Analysis of Consumer Preference

Kruskal - Wallis test: It is used to analyze whether significant differences exist in consumers' preferences for commodity categories at different ages. For our analysis, we will focus on the purchase frequency of food, daily necessities, and clothing among different age groups. We will calculate the Kruskal-Wallis statistics and P-value to determine whether the differences in purchasing behavior across these age groups are statistically significant. The results show that the P value is less than 0.05 in the food category, which indicates that there are significant differences in consumers' preferences for food in different age groups. As shown in Table 1:

Table 1 Consumer preference

Commodity category	Kruskal - Wallis statistic	p value	Is it significant ($\alpha = 0.05$)
Food	12.56	0.02	YES
Daily necessities	5.23	0.15	NO
Clothing	8.78	0.06	NO

Mann - Whitney U test: This study compares male and female consumers' preferences for specific commodity brands. Select a well-known cosmetics brand and analyze the purchase amount of male and female consumers. Calculate Mann-Whitney U statistics and p-value. The results show that the P value is less than 0.01, which demonstrates that there is a significant difference between male and female consumers' preferences for this cosmetic brand.

4.2.3 Cluster Analysis

In this study, non-parametric clustering methodologies (e.g., hierarchical clustering) are employed for the classification of consumers. The classification of consumers is predicated on their unique characteristics in terms of purchase behavior and preferences, which are then used to delineate distinct groups. The former category, designated as "loyal high-end customer groups," is characterized by elevated consumption frequency and substantial expenditure. In contrast, the latter category, termed "potential customer groups," is distinguished by diminished consumption frequency and minimal expenditure. Cluster analysis is helpful for us to formulate targeted marketing strategies for different groups.

4.3 Evaluation of Precision Marketing Effect

4.3.1 Marketing Activity Design

According to consumer preference analysis and clustering results, precision marketing activities are formulated for different groups. For "loyal high-end customer groups," we introduce exclusive member discounts and limited commodity preemption. For "potential customer groups," we send coupons and introduce new user exclusive activities, etc.

4.3.2 Evaluation Indicators and Methods

We selected sales, customer satisfaction, and customer repurchase rate as indicators to evaluate the effect of precision marketing. Relevant data were collected before and after the implementation of marketing activities, respectively, and marketing effects were evaluated through comparative analysis.

4.3.3 Evaluation Results

Table 2 Evaluation results

Evaluation indicators	Before the activity	After the activity	Rate of change
Sales (ten thousand yuan)	500	650	30%
Customer satisfaction (%)	80	85	6.25%
Customer repurchase rate (%)	30	35	16.67%

It can be seen from Table 2 that precision marketing activities have achieved remarkable results. Sales increased by 30%, customer satisfaction increased by 6.25%, and customer repurchase rate increased by 16.67%. It indicates that consumer preference analysis based on nonparametric statistical method can provide practical support for precision marketing under the influence of new retail business model, and assist enterprises improve marketing efficiency and economic benefits.

To sum up, nonparametric statistical methods have important application value in consumer preference analysis and precision marketing under the new retail business model. By rationally using these methods, enterprises will understand consumers' needs and preferences, formulate more accurate marketing strategies, and, as a result, gain an advantage in fierce market competition.

5. Conclusion and Prospect

This paper discusses the application of nonparametric statistical methods in analyzing consumer preferences and precision marketing in a new retail business model. It shows the great potential of these methods in dealing with unstructured data and data that do not follow specific distribution. Nonparametric statistical methods reduce assumptions about data structure and provide a flexible and robust data analysis approach, which is particularly valuable for examining the dynamic and complex data of the new retail sector. A more accurate market understanding can be obtained by applying nonparametric methodologies in the analysis of consumers' fundamental characteristics, purchasing behavior, product preference, marketing activity response, and dynamic changes in preference. These analysis results will help enterprises manage customer relationships, design personalized products and services, and formulate efficient marketing strategies.

As we look to the future, advancements in big data technology and the diverse methods of acquiring consumer data will enable more detailed and dynamic analysis of consumer preferences under the new retail business model. In applying nonparametric statistical methods, the researchers need to adapt to new challenges due to the development of big data, such as high-speed liquidity of processing data, high expansion of dimensions, and real-time updating of information. In addition, finding a way to better integrate nonparametric statistical methods into an intelligent business decision-making system to realize automation and immediate feedback in actual business operations will become an important direction of follow-up research and practice.

References

[1] Mielniczuk J. Nonparametric Statistical Inference with an Emphasis on Information-Theoretic Methods[J]. Entropy (Basel, Switzerland), 2022, 24(4).

[2] Bessa R J, Miranda V, Botterud A, et al. Time Adaptive Conditional Kernel Density Estimation for Wind Power Forecasting[J]. IEEE Transactions on Sustainable Energy, 2012, 3(4):660-669.

[3] Barros R S M D, Juan Isidro González Hidalgo, Cabral D R D L .Wilcoxon Rank Sum Test Drift Detector[J]. Elsevier, 2018.

[4] Chechile R A. A Bayesian analysis for the Mann-Whitney statistic[J]. Communications in Statistics - Theory and Methods, 2020, 49.

[5] Li X, Zhao X, Xu W A, et al. Measuring ease of use of mobile applications in e-commerce retailing from the perspective of consumer online shopping behaviour patterns[J]. Journal of Retailing and Consumer Services, 2020, 55.

[6] Vo MAN, Tran VAT, Ung-Pham T, Varela P, Nguyen QC. Combining text mining of social media data and conjoint approach to investigate consumer choices on organic food[J]. Food Quality and Preference, 2025, 124: 105369. DOI: 10.1016/j.foodqual.2024.105369.

[7] Etn V, Yildiz O. A comprehensive review on data preprocessing techniques in data analysis[J]. Pamukkale University Journal of Engineering Sciences, 2022, 28(2).