Sensitivity Analysis of Highway Customers with Different Vehicle Types to Discounts

Aijun Liu*[a]

[a]Inner Mongolia Autonomous Region Comprehensive Transportation Operation Coordination Center, 68 South Street, Geological Bureau, Saihan District, Hohhot, China

*laj@nmjt.com.cn; phone +86 1 384 818 0008

Abstract. Tolls are the direct income of highway operating companies, so how to effectively incentivize car owners to consume highway services is a highly concerned issue for the operating department. Providing discounts (such as discounts and red envelopes) to car owners is a commonly used method in highway operation. However, inclusive discounts cannot maximize the revenue of highway companies, as users have varying sensitivity to discounts. Some users, regardless of whether they have discounts or not, will take to the highway. Therefore, it is necessary to identify these customers. This paper uses correlation analysis method to identify the sensitivity of users of different vehicle types to discounts. This method provides a basis for formulating discount strategies for highway operations, thereby maximizing the savings in discount costs and improving operational revenue.

Keywords: Sensitivity, highways, toll, preferential, correlation coefficient

1 INTRODUCTION

In the past thirty years, China's road infrastructure construction has made rapid progress. The construction of highways adopts the strategy of "building road with loan and repaying loan with toll", which makes it necessary for highway operations to be profitable in order to repay bank loans on time. In recent years, the highway freight market has been plagued by a downturn, especially in the first half of 2023 when it was exposed more frequently. Various provinces have successively announced differentiated pricing policies for toll roads, implementing differentiated pricing policies based on road sections, time periods, vehicle types, and entrances and exits.

Highway companies achieve operational benefits by collecting tolls, and the amount of tolls is related to many factors. For example, the discount rate is one of the important factors. When the discount rate is large, more car owners will choose to travel on the highway. However, when there are too many vehicles on the highway, it will reduce traffic efficiency, leading to a decrease in revenue. Reasonably adjusting the discount range and scope, and maintaining highway vehicles in an optimal state, is an effective method to achieve maximum revenue. The prerequisite for achieving this goal is to discover the inherent relationship between tolls and discount rates.

B. K. Kandel et al. (eds.), *Proceedings of the 2023 8th International Conference on Engineering Management (ICEM 2023)*, Atlantis Highlights in Engineering 23,
https://doi.org/10.2991/978-94-6463-308-5_42

2 Related Work

Price reduction promotion is a useful strategy for many businesses to increase store sales. If the price promotion framework adopts single discount/multiple discounts - first large then small/multiple discounts - first small then large, etc., the promotion framework will have an impact on consumers' purchase intention [1]. There is evidence to suggest that some consumers' reactions to promotional signals do not take into account relative price information. Regardless of whether the price of the promotional brand decreases or not, individuals with low cognitive demand will respond to the simple existence of promotional signals, but individuals with high cognitive demand will only respond to promotional signals when they are accompanied by substantial price reductions [2]. Under scarcity conditions, consumers' perceptions of quality and financial sacrifice exhibit different response patterns, depending on relative price levels and consumers' motivation to process information [3].

Correlation analysis refers to the analysis of two or more correlated variable elements, in order to measure the degree of correlation between two variable factors. There needs to be a certain connection or probability between the elements of correlation in order to conduct correlation analysis. Correlation is not equivalent to causality, nor is it simply personalization. The scope and fields covered by correlation almost cover all aspects we see, and the definition of correlation varies greatly among different disciplines. Correlation coefficient is an important method for measuring data correlation, mainly including Pearson, Spearman, and Kendall correlation coefficients [4].

This paper discusses the relationship between highway tolls and discounts, mainly using correlation coefficients to reflect the sensitivity of car owners to optimizing highway prices.

3 Calculation Method for Sensitivity of Highway Discounts

3.1 Basic ideas

Based on the differences in correlation coefficients among different vehicle types, targeted preferential policies can be formulated. Although tolls are also related to factors such as time periods and weather, this paper only analyzes the relationship between tolls and discounts from the perspective of vehicle types.

In China, highways on statutory holidays are generally free of charge, with a high traffic volume and zero toll revenue. It is not possible to discover the inherent relationship between toll and discounts. On weekdays, there is less toll traffic on the highway, so traffic data can reflect the relationship between the two. In response to the above issues, traffic data for holidays has been removed and weekday traffic data has been retained.

Collect ETC traffic data, calculate Pearson, Spearman, and Kendall correlation coefficients between tolls and discounts, and use the correlation coefficients as the sensitivity of car owners to discounts. Due to the differences in correlation coefficients among different vehicle types, different vehicle types are distinguished in the calculation process of correlation coefficients. In other words, sensitivity is the correlation coefficient between tolls and discounts for different car types.

3.2 Data selection and standardization processing

Data preprocessing is an important step and the first step in machine learning modeling process [5]. In the application fields where datasets with tens or hundreds of thousands of variables are available, variable and feature selection has become the focus of many studies. The goal of variable selection is to improve the predictive performance of predictive factors, provide faster and more cost-effective predictive factors, and better understand the basic process of generating data [6, 7, 8].

Aggregation Sampling

Due to the fact that the minimum time granularity for collecting toll sample data is seconds, the sample size is large, especially for highways in some central cities with huge traffic volume. Therefore, after excluding holiday samples, further processing is still needed. This paper adopts the method of time aggregation for sampling, with the basic idea of aggregating by vehicle type to hourly granularity, and summarizing hourly tolls and discounts as one sample data.

The specific projects are as follows:

(1) The selected descriptive features include: date, time period, and vehicle type;

(2) The relevant calculation features are: tolls, discount limits;

(3) Granularity: the time period is hours.

Normalization

Data standardization is a commonly used method in machine learning. If the original indicator values are directly used for analysis, the role of indicators with higher numerical values in comprehensive analysis will be highlighted, and the role of indicators with lower numerical levels will be relatively weakened. Data standardization can avoid numerical problems, such as numerical problems caused by large numbers. At the same time, it can also balance the contributions of various features [9]. Common data standardization techniques include normalization, standardization, discretization, etc. [10].

This paper adopts the Min-Max Normalization method, which compresses the data of a feature X into the range of [0,1]. The specific calculation formula is

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

3.3 Correlation calculation

The first issue to consider in correlation analysis is whether there may be a correlation between the two variables. If a positive conclusion is obtained, then it is necessary to proceed with the next quantitative analysis. Through investigation, it was found that there is a correlation between toll revenue and discounts on highways, so correlation coefficients can be used to quantitatively analyze the correlation.

This paper uses Pearson, Spearman, and Kendall correlation coefficients to analyze the correlation between tolls and discounts from different dimensions. The three correlation coefficients are all measures of the degree of correlation between variables, and due to their different calculation methods, their purposes and characteristics are also different.

The Pearson correlation coefficient is calculated based on the variance and covariance of the original data, so it is sensitive to outliers and measures linear correlation. Therefore, even if the Pearson correlation coefficient is 0, it can only indicate that there is no linear correlation between variables, but there is still a possibility of curve correlation.

The Spearman and Kendall correlation coefficients are both based on the relative size of rank and observations, and are a more general non parametric method that is less sensitive to outliers and therefore more tolerant. They mainly measure the connections between variables.

(1) Pearson Correlation Coefficient

Used to measure the linear correlation between two variables X and Y. It has values between+1 and -1, where 1 is the total positive linear correlation, 0 is the nonlinear correlation, and -1 is the total negative linear correlation. A key mathematical characteristic of the Pearson correlation coefficient is that it remains constant under individual changes in the position and scale of two variables. That is to say, transform X to $a + bX$ and Y to $c + dY$ without changing the correlation coefficient, where a, b, c, and d are constants and $b, d > 0$.

The calculation formula corresponding to Pearson correlation coefficient is

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{2}$$

(2) Spearman Correlation Coefficient

Spearman uses the rank size of two variables for linear correlation analysis, without requiring the distribution of the original variable, and belongs to non-parametric statistical methods. Therefore, its applicability is much wider than the Pearson correlation coefficient. For data that follows the Pearson correlation coefficient, the Spearman correlation coefficient can also be calculated, but the statistical efficiency is lower than the Pearson correlation coefficient (it is not easy to detect the actual correlation between the two).

If there are no duplicate values in the data and two variables are completely monotonically correlated, the Spearman correlation coefficient is+1 or -1. Even if there are outliers in the Spearman correlation coefficient, the influence on the Spearman correlation coefficient is very small because the rank of the outliers usually does not change significantly.

For samples with a sample size of n, n raw data are converted into hierarchical data, and the Spearman correlation coefficient ρ by

$$\rho = \frac{\sum_{i=0}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=0}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=0}^{n}(y_i - \overline{y})^2}} \tag{3}$$

In practical applications, the connections between variables are irrelevant, so they can be calculated through simple steps ρ. The difference between the levels of two observed variables, then ρ by

$$\rho = 1 - \frac{6 \cdot \sum d_i^2}{n(n^2 - 1)} \tag{4}$$

(3) Kendall Correlation Coefficient

The Kendall correlation coefficient is also a rank correlation coefficient used as an indicator to reflect the correlation of categorical variables. It is suitable for situations where both variables are classified in an orderly manner, using the Greek alphabet τ Represents its value.

The value range of τ is between -1 and 1:

- When τ When it is 1, it indicates that two random variables have consistent hierarchical correlation;
- When τ When it is -1, it indicates that two random variables have completely opposite hierarchical correlations;
- When τ When it is 0, it indicates that two random variables are independent of each other.

The Kendall coefficient is based on the idea of synergy. For the two pairs of observations of X and Y, X_i, Y_i, and X_j, Y_j, if $X_i < Y_i$ and $X_j < Y_j$, or $X_i > Y_i$ and $X_j > Y_j$, then these two pairs of observations are considered harmonious, otherwise they are considered disharmonious. The calculation formula for Kendall's correlation coefficient is

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of disconcordant pairs})}{\frac{1}{2}n(n-1)} \tag{5}$$

Subtract the number of harmonious observation pairs (consistent observation pairs) from the number of disharmonious observation pairs (inconsistent observation pairs), divided by the total logarithm of observation values.

4 Empirical Analysis

4.1 Data collection

We selected toll data from a certain city in China for June 2022, with a minimum time granularity of seconds, a sample size of 826424, and a feature count of 70. Select three characteristics: vehicle type, discount amount, and toll, and use formula (1) to normalize the discount amount and toll. Figure 1 shows the distribution of toll and discount data without considering specific vehicle types. Due to the large sample size, 200 samples were randomly selected for display.

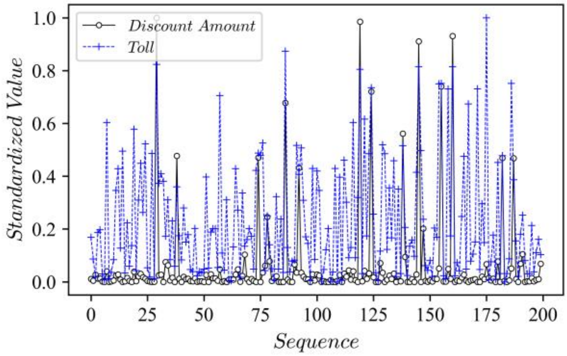


Fig.1. Distribution of sample data without considering vehicle types tolls and discounts

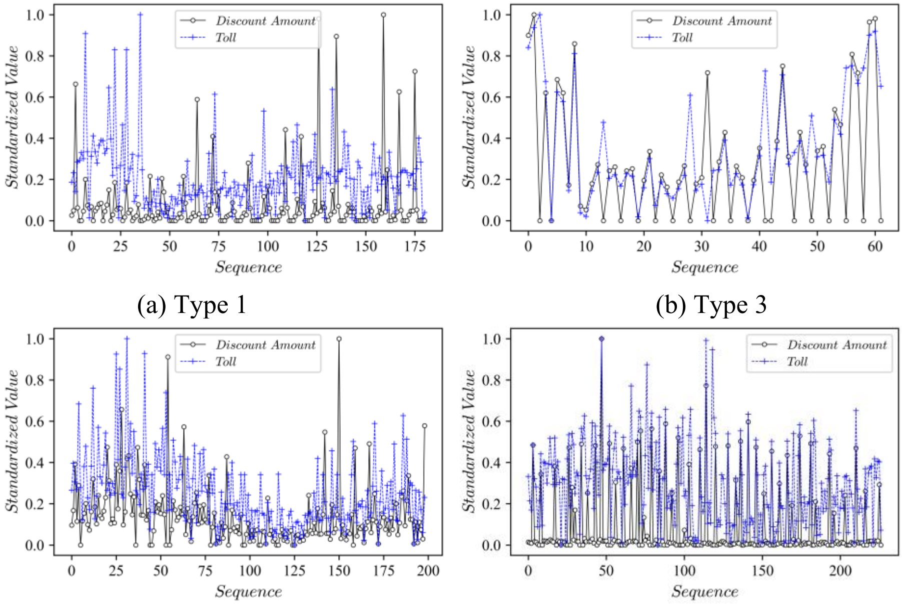
4.2 Vehicle type data

The sample dataset contains vehicle type data. Due to the different sensitivity of passenger cars and trucks in the same category to discounts, different vehicle types were subdivided and coded, as shown in Table 1. The definition section in Table 1 describes the main characteristics of the vehicle model.

Table 1. Vehicle Type Code and Definition

Type	Code	Definition
Type 1	1	Sedan and small passenger cars with less than 7 seats (including 7 seats)
	11	Small trucks of less than 2 tons (including 2 tons)
Type 2	2	8-19 seat (including 19 seats) passenger cars
	12	2-5 ton (including 5 ton) freight car
Type 3	3	20-39 seater (including 39 seater) passenger cars
	13	5-10 ton (including 10 ton) truck
Type 4	4	40 or more passenger cars
	14	10-15 ton (including 15 ton) truck and 20 foot container truck
Type 5	15	15 ton and above trucks
	16	40 foot container truck

The distribution of discount amounts and tolls varies for different car types. As shown in Figure 2, Figure 2 (a) shows the distribution of the discount amount (dashed part) and toll (solid part) for Type 1. Intuitively, the correlation between the two is smaller than Type 3 in Figure 2 (b), Type 13 in Figure 2 (c), and Type 16 in Figure 2 (d).



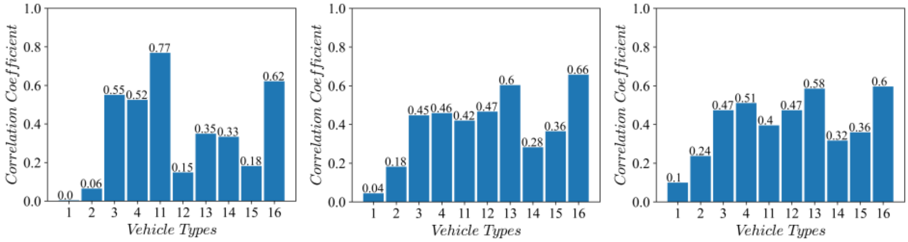
(a) Type 1 (b) Type 3

(c) Type 13 (d) Type 16

Fig. 2. Distribution of Discount Quota and Toll for Different Vehicle Types

4.3 Empirical results

Calculate the Pearson, Spearman, and Kendall correlation coefficients for different vehicle types using formulas (2), (4), and (5). The results are as shown in Figure 3.



(a) Pearson (b) Spearman (c) Kendall

Fig. 3. Correlation Coefficients for Different Vehicle Types

The toll and discount amount for different car types are positively correlated. For the Person correlation coefficient, types 3, 4, 11, and 16 are all at a high level, as shown in Figure 3 (a). For Spearman and Kendall correlation coefficients, types 3, 4, 11, 12, 13, 14, 15, and 16 are all at a high level. The three correlation coefficients basically maintain the same distribution.

Therefore, the conclusions are as following:

(1) 8-19 seat (including 19 seats) or less passenger cars are not sensitive to discounts;

(2) All trucks are sensitive to discounts;

(3) Buses with 20 to 39 seats (including 39 seats) and above are sensitive to discounts.

According to the analysis of other characteristic data of the corresponding vehicle types, it can be seen that the majority of types 1 and 2 are private cars, belonging to middle-class families, with a majority of short trips, and can bear highway tolls. The vast majority of trucks are operating vehicles, and there are many mediums to long distance vehicles. Differential pricing makes car owners more concerned about costs, so they are sensitive to discounts. Similarly, buses with 20-39 seats (including 39 seats) and above are mostly tourist buses, and cost control is also a concern.

5 Conclusion

Tolls are a direct benefit for highway operating companies and are closely related to preferential policies. This paper proposes a method for calculating the correlation between tolls and discounts for vehicle types. The tolls and discounts are mapped to the same space [0,1], and the vehicle types are encoded. The Pearson, Spearman, and Kendall correlation coefficients of the two are calculated to measure the sensitivity of different vehicle types to discounts. An empirical study was conducted on a dataset of highway tolls in a certain city in China, and the results showed that private small buses are not sensitive to discounts, while operating trucks and buses are sensitive to discounts. This result provides a basis for the highway operation department to develop differentiated preferential strategies.

References

1. Wadhwa, M., Zhang, Y., Dass, M. The effect of discount framing on consumer purchase decisions: A meta-analysis. Journal of Retailing and Consumer Services, 40, 221-230(2018).
2. Inman, J. J., McAlister, L., Hoyer, W. D. Promotion signal: Proxy for a price cut? Journal of Consumer Research, 17(1), 74-81(1990).
3. Nunes, J. C., Park, C. W. The effects of perceived scarcity on consumers' processing of price information. Journal of Consumer Research, 29(3), 320-334(2003).
4. M.M Mukaka, A guide to appropriate use of Correlation coefficient in medical research, Malawi Medical Journal; 24(3), 69-71(2012).
5. García, S., Luengo, J., Herrera, F. Data preprocessing in data mining. Springer, (2010).
6. Guyon, I., Elisseeff, A. An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182(2003).
7. Peng, H., Long, F., Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 1226-1238(2005).
8. Yu, L., Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), 856-863(2003).

9. Batista, G. E., Prati, R. C., Monard, M. C. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter, 6(1), 20-29(2004).
10. Guyon, I., Elisseeff, A. An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182(2003).