Erzincan Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
2023, 16(2), 598-607
ISSN: 1307-9085, e-ISSN: 2149-4584
Araştırma Makalesi

Erzincan University
Journal of Science and Technology
2023, 16(2), 598-607
DOI: 10.18185/erzifbed.1113137
Research Article

Comparison of passenger wait and travel time perceptions at bus stops

Buket Çapalı [1*]

[1]Suleyman Demirel University, Faculty of Architecture, Urban and Regional Planning, Isparta

Received: 06/05/2023, **Revised:** 23/08/2023, **Accepted:** 23/08/2023, **Published:** 31/08/2023

Abstract

Passenger's perceived time is essential in improving the public transport system. Since it is only possible to reach these values with a survey, it is complicated to obtain them. Therefore, studies on this subject are very few and valuable. In the study, passengers' perceived waiting time and travel time values were determined at bus stops. To obtain these values, a survey was conducted with the passengers. The linear relationship between waiting time and travel time data was examined, and a model was created. The data obtained with the basic regression model, which is the basis of machine learning, is divided into 20% test and 80% training data. The estimation was made with 85% success with the tested data.

Keywords: Urban transportation planning; passenger wait time perceptions; passenger travel time perceptions; poisson ratio; ordinary least square method; linear regression

Otobüs duraklarında yolcu bekleme ve seyahat süresi algılarının karşılaştırılması

Öz

Yolcunun algıladığı zaman, toplu taşıma sisteminin iyileştirilmesinde önemlidir. Bu değerlere ancak anket ile ulaşmak mümkün olduğu için elde etmek oldukça zordur. Bu nedenle bu konudaki çalışmalar çok az ve değerlidir. Çalışmada yolcuların otobüs duraklarında algıladıkları bekleme süreleri ve seyahat süreleri değerleri belirlenmiştir. Bu değerleri elde etmek için yolcularla anket çalışması yapılmıştır. Bekleme süresi ile seyahat süresi verileri arasındaki doğrusal ilişki incelenmiş ve bir model oluşturulmuştur. Makine öğrenmesinin temeli olan temel regresyon modeli ile elde edilen veriler %20 test ve %80 eğitim verisi olarak ikiye ayrılmaktadır. Test edilen verilerle %85 başarı ile tahmin yapılmıştır.

Anahtar Kelimeler: Kent içi ulaşım planlaması; yolcu bekleme süresi algısı; yolcu seyahat süresi algısı; poison oranı; ki-kare metodu; lineer regresyon

*Corresponding Author: buketcapali@sdu.edu.tr
Buket Çapalı, https://orcid.org/0000-0003-1917-1654

1. Introduction

Travel time is one of the most important dimensions of the level of service that passengers perceive. In particular, the total travel time is usually divided into in-vehicle and out-of-vehicle time for benefit function-based modeling. The waiting time, an important part of the time out of the vehicle, is more effective than the time in the vehicle and is often shown as one of the most important factors affecting the transport type selection [1].

The usual approach to waiting time in public transport is that the estimated waiting time is half the frequency of the transport vehicles [2]. This approach is based on the assumption that passengers arrive at the station haphazardly, passengers aboard the first vehicles, and the service is reliable. Another study found that providing time information in their study has affected the detected wait time for more than 15 minutes. In the studies, it has been determined that the saving made from the waiting time is not of great economic importance [3]. However, it is an important factor in terms of the satisfaction of public transportation users due to its psychological effect [4].

In an article investigating the relationship between actual and perceived waiting times in public transport, he found that passengers had a greater estimate of the average waiting time than 14%. The study found that the anticipation of the waiting time for short-term passengers is higher than those waiting for a long time [5].

Another study modeled the waiting time by considering the physical conditions of the line and the station and the effect of reliability and passenger information systems. As a result of the linear models they created, they saw that the presence of the information system shortens the waiting time by an average of 2 minutes and reliability by an average of 1.3 minutes [6].

According to an article are two ways available to manage the waiting process. The first is to reduce the actual length of the wait (for example, increase bus frequency) with business management techniques. As a second method, it is suggested that the perceived waiting time can be reduced with the help of some psychological effects and this can be as effective as reducing the waiting time [7].

System visuals showing the arrival time of the next vehicle at stops and stations greatly reduce anxiety. Only such systems can build trust for all public transport systems and improve the system's image. It is accepted that the perceived reliability at the stops is positively affected by the new visuals used. Therefore, the service is perceived as more reliable. Thus, the perceived service level increases [8].

Studies investigating the approaches and perceptions of passengers are given great importance. It was concluded that passenger information systems generally positively affect passengers' waiting time expectations [9]. It has been shown that such systems reduce the perceived waiting time by 26% [10]. Discussions in the literature suggest that perceived and measured benefits vary depending on where the research is conducted. This study examines the relationship between the passengers' waiting time and travel time responses at the surveyed stop with statistical models.

2. Material and Methods

The waiting time, the individual time measured and spent, is the actual waiting time if measured with a stopwatch, and the perceived waiting time if it is obtained with the help of a survey. This study examined the linear relationship between passengers' perceptions of waiting time and travel time. The survey determined the waiting time and travel time detected in the study. Therefore, the data in the study show the perceived durations.

The methodology includes Ordinary Least Square (OLS), poisson, linear regression (LR). First of all, a linear model will be created, which is used to describe the relationship between two variables, by looking at the correlation between the data. Simple linear regression happens when an independent variable is used to predict a dependent variable. In simple linear regression, there are two variables: the dependent variable and the independent variable. The independent variable is the waiting time, the dependent variable is the travel time.

2.1. Ordinary Least Square (OLS), Poisson and Linear Regression

For simple linear regression it is important to apply the ordinary least squares method (OLS). Simple linear regression is a model that shows the relationship between the dependent variable x and the independent variable y. The model relation is given in equation 1.

$$y = \beta 0 + \beta 1x + \varepsilon \quad (1)$$

Where $\beta 0$ is intercept, $\beta 1x$ is slope (unkown constant) is ε random error component. The OLS method is used to estimate β0 and β1. The OLS method seeks to minimize the sum of the squared residuals. This means from the given data we calculate the distance from each data point to the regression line, square it, and the sum of all of the squared errors together.

Ordinary Least Squares method, often called linear regression, is a supervised machine learning algorithm used for parameter estimation of different functional relationships [11]. The Poisson ratio, which was first calculated by Sime´on Denis Poisson (1781-1840), was proposed as an elastic constant by Pearson (1886) in the following years [12].

In this study, the least squares method is used to explain the correlation between two variables and its effect on parameter values. Linear regression with OLS and performance of each explanatory variable, appropriate values of predicted parameters and alignment coefficients were determined. The order of the variables in the regression model was determined by calculating the values of the correlation coefficients between the dependent and independent variables. The ratio between the minimum absolute value of the correlation coefficient between the dependent and independent variables and the maximum absolute value of the correlation coefficient was calculated. The absolute values of the correlation coefficients between the dependent and independent variables and the degree of linearity caused by each variable in the regression model were determined.

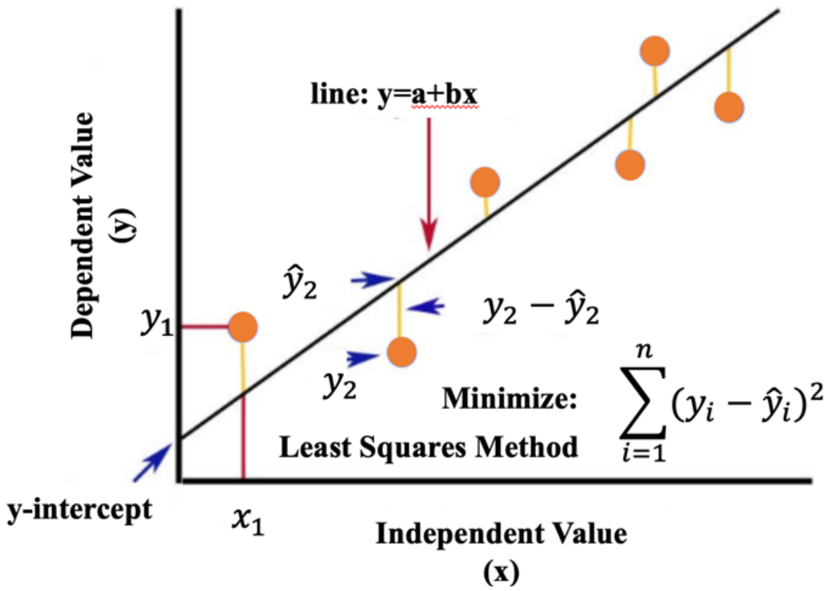

Figure 1. Least square method between dependent and independent value

3. Results and Discussion

The survey data is loaded into the "df" variable. The description of the dataset is also written for a better dataset insight. The standard deviation value is used to determine whether the data is consistent. The standard deviation values, which are the central distribution scale, are found in Table 1.

Table 1. Statistical descriptions of dependent and independent variables

	Waiting time	*Travel time*
count	506.000000	506.000000
std	4.933021	5.052719
min	5.000000	5.000000
25%	15.000000	15.000000
50%	15.000000	20.000000
75%	20.000000	24.000000
max	30.000000	35.000000

The standard deviation of the waiting time is 4.93 and the standard deviation of the travel time is 5.05. Values indicate that the data is consistent. The minimum value of both values is 5. The first quartile of both values is the same. The second quartile value also gives the median value.

The difference between waiting time and travel time is 5 minutes. This difference has decreased to 4 minutes at the third quartile value. Table 1 is obtained with the code “df.describe()”.

The significance of the mean of the two variables was determined by covariance analysis. At this stage, pre-test measurements are defined as covariates. Table 2 is obtained with the code “df.cov()”. covariance is the relationship between two variables. Table 2 shows that there is a positive relationship. Since the covariance is affected by the magnitude of the measurement values, it cannot give a standard value.

Table 2. Covariance of dependent and independent variables

	Waiting time	***Travel time***
Waiting time	24.334697	20.166552
Travel time	20.166552	25.529969

Table 3 is obtained with the code “df.corr()”. Since the covariance is affected by the magnitude of the measurement values, we need to obtain the correlation values. Correlation offers a value between +1 and -1. Travel time and waiting time correlation coefficient of 0.8 indicates that the relationship between them is high.

Table 3. Correlation of dependent and independent variables

	Waiting time	***Travel time***
Waiting time	1.000000	0.809084
Travel time	0.809084	1.000000

In order to obtain the "most accurate" regression model, the Ordinary least squares (OLS) method and the Poisson method were applied first. In this study, OLS was used to find a linear model to predict the relationship between travel time and waiting time in the dataset obtained from the survey.

Simple linear regression was calculated using the "ordinary least squares" method to discover the most suitable row for the travel time and waiting time matched dataset obtained according to the passengers' perceptions.

```
                        Results: Ordinary least squares
==============================================================================
Model:                  OLS              Adj. R-squared (uncentered): 0.973
Dependent Variable:     travel_time      AIC:                         2638.1031
Date:                   2022-03-14 17:41 BIC:                         2642.3296
No. Observations:       506              Log-Likelihood:              -1318.1
Df Model:               1                F-statistic:                 1.803e+04
Df Residuals:           505              Prob (F-statistic):          0.00
R-squared (uncentered): 0.973            Scale:                       10.738
----------------------------------------------------------------------------------
                 Coef.      Std.Err.        t         P>|t|       [0.025     0.975]
----------------------------------------------------------------------------------
waiting_time     1.0987       0.0082     134.2771     0.0000      1.0826     1.1147
------------------------------------------------------------------------------
Omnibus:                   1.984            Durbin-Watson:               1.845
Prob(Omnibus):             0.371            Jarque-Bera (JB):            1.743
Skew:                      -0.002           Prob(JB):                    0.418
Kurtosis:                  2.712            Condition No.:               1
==============================================================================
```

Figure 2. OLS of dependent and independent variables

With OLS the most important decision is which features to use in prediction and how to use them. "Linear" means linear in coefficients only; these models can handle many kinds of functions (Figure 2). Many approaches exist for deciding which features to include. For now we will only use cross-validation. The Poisson distribution was used to determine and compare the weight of the independent variable (Figure 3).

```
Optimization terminated successfully.
         Current function value: 5.696245
         Iterations 30
                         Results: Poisson
================================================================
Model:              Poisson          Pseudo R-squared: -0.856
Dependent Variable: travel_time      AIC:              5766.6000
Date:               2022-03-14 17:41 BIC:              5770.8265
No. Observations:   506              Log-Likelihood:   -2882.3
Df Model:           0                LL-Null:          -1553.1
Df Residuals:       505              LLR p-value:      nan
Converged:          1.0000           Scale:            1.0000
No. Iterations:     30.0000
-----------------------------------------------------------------
               Coef.   Std.Err.     z      P>|z|   [0.025  0.975]
-----------------------------------------------------------------
waiting_time   0.1478    0.0005  291.9802  0.0000  0.1468  0.1488
================================================================
```

Figure 3. Poisson of dependent and independent variables

First of all, it is necessary to determine the correct coefficients of the model. After these coefficients are determined, the model should predict the target variable value without making any mistakes. After these stages, the linear regression model can be determined with a low error.

In this study, linear regression analysis was used to estimate the value of the travel time variable relative to the value of the waiting time variable. The variable to be estimated was called the dependent variable. The variable used to estimate the value of the other variable was called the independent variable.

Figure 4 shows the graph of the relationship between waiting time and travel time. After the values were normalized, the graph was obtained.

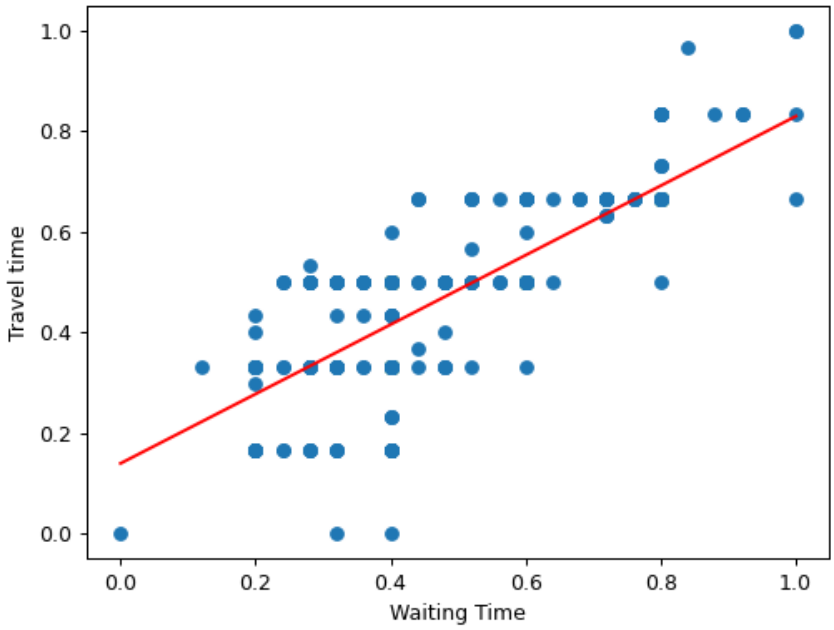

Figure 4. Relationship between waiting and travel time

Placing a linear model in a dataset is the simplest procedure. To evaluate the performance of the model, we divide the dataset into a training and test set, and then compare the model's predictions on the test set with the performance of the trained model. If the model has a high Mean Suarred Error (MSE) in both the training and test set, it is insufficient. It is overfitting if it has a small MSE in the training set and a high MSE in the test set. In the study, while supported by the LinearRegression object in scikit-learn, the MSE value of 0.0988 was calculated with the average_squared_error() function. In the study, the r^2 value was calculated as 0.85. With the linear regression analysis method, the coefficients of the linear equation were estimated by using the waiting time variable that best predicted the value of the travel time variable. The regression model relation established between travel time and waiting time is given in equation 2.

$$y = 0.14 + 0.69x \tag{2}$$

The relationship between travel time and waiting time variables was modeled using linear regression. For the linear model, divide the dataset into the training and test dataset. Then the Linear Regression Model is trained with the training data. After obtaining the prediction result, the mean squared error was evaluated for the training dataset and then compared with its value in the test dataset prediction to see if the value was comparable. Graph was obtained using pyplot in matplotlib.

When Figure 5 is examined, the normalized values obtained and estimated are seen. It is seen that the estimated values are close to the average of the obtained values and are consistent.

There are 506 values. The orange color represents the predicted values and the blue color represents the obtained values.

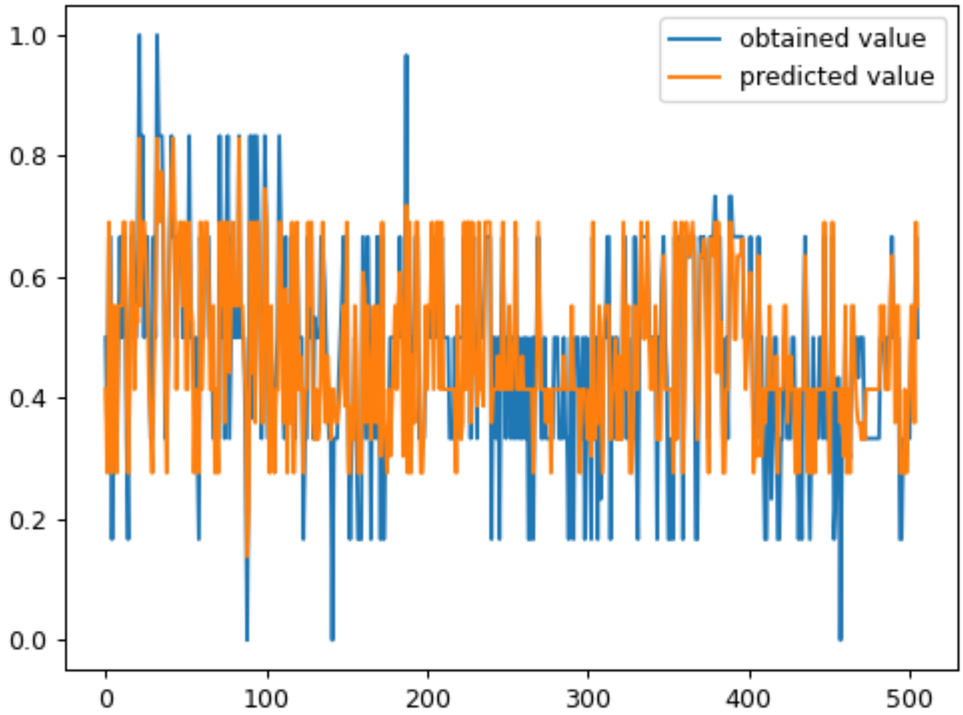

Figure 5. Obtained and predicted value

4. Conclusion

This study aims to establish a correlation between travel time and waiting time for urban public transportation. It is very difficult to estimate the travel time. Because many different variables affect the travel time. The analysis in this study is based on the survey responses of 506 passengers using a public transport line. Travel time and waiting time data of passengers are obtained in the survey. The linear regression estimation algorithm method is used to determine the relationship between the travel time and the answers given by the users to the waiting time questions from the data obtained. One of the prediction algorithms of supervised machine learning methods is the linear regression algorithm. Linear regression is one of the most well-known statistics and machine learning algorithms. In this study, the poisson and chi-square distributions, which are highly preferred in inferential statistical analysis, were used.

When the results obtained were examined, it was seen that the perception of passenger waiting was high. The waiting time perception should be half of the bus frequency value. But it's almost twice the value on average. This is because the selected line is crowded, so the bus passes without stopping at the stop for the passengers. Therefore, the passenger has to wait for the next bus. In this case, it reduces the passenger's confidence in the public transport system. For this reason, it is necessary to reduce the passenger's waiting time with the express line application to the crowded stops of the line.

Ethics in Publishing

There are no ethical issues regarding the publication of this study

Acknowledgement

I would like to thank the Municipality of Isparta and Private Public Buses Operator for allowing the survey.

Author Contributions

Capali B. wrote the article, read and approved the final version.

Conflict of Interests

The author declares that there is no conflict of interest.

References

[1] Durrande-Moreau, A. (1999) Waiting for service: ten years of empirical research. International Journal of Service Industry Management, 10(2), 171-194.

[2] Hall, R. W. (2001) Passenger waiting time and information acquisition using automatic vehicle location for verification. Transportation Planning and Technology, 24(3), 249-269.

[3] Hui, M. K., Tse, D. K. (1996) What to tell consumers in waits of different lengths: An integrative model of service evaluation. Journal of Marketing, 60(2), 81-90.

[4] Wardman, M., Hine, J., Stradling, S. G. (2001) Interchange and Travel Choice, Vol. 1, Vol. 2 and Research Summary. Edinburgh: Central Research Unit, Scottish Executive.

[5] Moreau, A. (1992) Public transport waiting times as experienced by customers. Public transport international, 41(3).

[6] Doğan, G., Özuysal, M. (2017) Toplu ulaşımda bekleme süresini etkileyen faktörlerin incelenmesi: güvenilirlik, yolcu bilgilendirme sistemi ve fiziksel koşullar. Teknik Dergi, 28(3), 7927-7954.

[7] Larson, R. C., Larson, B. M., Katz, K. L. (1991) Prescription for waiting–in line blues: Entertain, enlighten and engage. Sloan Management review,(winter), 32(2), 44-55.

[8] Infopolis2. (1998). Review of Current Passenger Information Systems.

[9] Forsyth, E., D. Silcock. (1985) "Real Time Information for Passengers on the London Underground. ." PTRC SAM Seminar J 157–71.

[10] Schweiger, C. L. (2003) Real-time bus arrival information systems (No. 48). Transportation Research Board.

[11] Pavelescu, F. M. (2004) Features of the ordinary least square (OLS) method. Implications for the estimation methodology. Journal for Economic Forecasting, 1(2), 85-101.

[12] Todhunter, I. (2014) A History of the Theory of Elasticity and of the Strength of Materials (Vol. 2). Cambridge University Press.