

# Analysis of passenger boarding time difference between adults and seniors based on smart card data

Lei Da Chen Xuewu Cheng Long Luo Ronggen

(School of Transportation, Southeast University, Nanjing 210096, China)

**Abstract:** As an essential component of bus dwelling time, passenger boarding time has a significant impact on bus running reliability and service quality. In order to understand the passengers' boarding process and mitigate passenger boarding time, a regression analysis framework is proposed to capture the difference and influential factors of boarding time for adult and elderly passengers based on smart card data from Changzhou. Boarding gap, the time difference between two consecutive smart card tapping records, is calculated to approximate passenger boarding time. Analysis of variance is applied to identify whether the difference in boarding time between adults and seniors is statistically significant. The multivariate regression modeling approach is implemented to analyze the influences of passenger types, marginal effects of each additional boarding passenger and bus floor types on the total boarding time at each stop. Results show that a constant difference exists in boarding time between adults and seniors even without considering the specific bus characteristics. The average passenger boarding time decreases when the number of passenger increases. The existence of two entrance steps delays the boarding process, especially for elderly passengers.

**Key words:** elderly passengers; smart card data; boarding time differences; analysis of variance; regression analysis; marginal effect

**DOI:** 10.3969/j.issn.1003-7985.2019.01.014

Public transit has been regarded as an essential solution to the increasing challenges of urban traffic congestion, especially in some developing countries like China. Governments and public transit agencies have taken several measures to make the public bus services more attractive to citizens.

Among them, one measure for attracting more passengers to take buses is to ensure bus service reliability, and another critical measure is to promote accessibility for some special passenger groups like seniors. The dwelling

time accounts for a substantial proportion of bus travel time, and thus it has a significant effect on service reliability. Besides, it is clear that the dwelling time is mainly spent in allowing passengers to get on and off the bus. Movements in the whole process of taking a bus, especially the boarding movement, reflect the accessibility to bus service for each type of passengers. Hence, a critical issue is to study the passenger boarding movement.

Previous studies mainly focused on investigating and analyzing the determinants of bus dwelling time using field survey data on a macro level, whereas the microscopic characteristics of bus passengers in the boarding or alighting process receive little attention<sup>[1-3]</sup>. Furth<sup>[4]</sup> found that higher variability of dwelling time might increase the risk of bus service unreliability. Based on the manually collected data, Tirachini<sup>[5]</sup> identified the effects of different factors, including the type of bus floor, the age of passengers as well as the type of fare collection system. Furthermore, by using the archived bus dispatch system data, Bertini and El-Geneidy<sup>[6]</sup> found that the dwelling time per passenger was not necessarily related to the total number of passengers. Shalaby et al.<sup>[7]</sup> proposed a bus travel time prediction model and discovered that the boarding passengers had a more significant influence than the alighting ones on dwelling time. Li et al.<sup>[8]</sup> developed a simulation model to estimate the bus dwell time, and similarly, González et al.<sup>[9]</sup> established a bus dwelling time estimation model which considered the influence of occasional incidents. Then the emergence of smart card technology provided a new and more efficient way for researchers to model the bus dwelling time and analyze the relationship between dwelling time and the boarding/alighting process. Using the smart card data, Widanapathirane et al.<sup>[10]</sup> presented a model of busway station dwelling time, while Sun et al.<sup>[11]</sup> used the smart card data for the first time to study the bus boarding and alighting dynamics by measuring the interval between two successive boarding/alighting passengers. Before their study, the average dwelling time per passenger was treated as a constant. However, they did not explore the distinct boarding/alighting characteristics of passengers at different ages. Moreover, Zhao et al.<sup>[12]</sup> introduced the index of rank to identify how passenger boarding/alighting time changed in a dynamic process.

Although some researchers applied the smart card data

**Received** 2018-07-20, **Revised** 2018-12-10.

**Biographies:** Lei Da (1993—), male, Ph. D. candidate; Chen Xuewu (corresponding author), female, doctor, professor, chenxuewu@seu.edu.cn.

**Foundation item:** The National Natural Science Foundation of China (No. 51338003, 71801041).

**Citation:** Lei Da, Chen Xuewu, Cheng Long, et al. Analysis of passenger boarding time difference between adults and seniors based on smart card data[J]. Journal of Southeast University (English Edition), 2019, 35(1): 97 – 102. DOI: 10.3969/j.issn.1003-7985.2019.01.014.

to the study of passenger boarding/alighting behavior, few of them took ages into account in their articles. However, this attribute of passengers plays a vital role in determining passenger movement time since different types of passengers may spend different amounts of time when boarding or alighting a bus.

In China, most of the smart card systems only have passenger boarding records. Given this limitation, our study focuses on the microscopic analysis of the passenger boarding process. There are two objectives in this study: the first one is to identify the difference of the boarding time between adults and seniors, and the second one is to investigate the effect of various factors on passenger boarding time.

## 1 Data Collection and Preprocessing

### 1.1 The smart card system in Changzhou

Changzhou is a prefecture-level city in southern Jiangsu province of China, bordering the provincial area of Nanjing to the west and Zhejiang province to the south. As a part of China's highly developed Yangtze Delta region, Changzhou now covers an area of 4 385 square kilometers, with a total population of 4.7 million in its five urban districts and one satellite city. There is no metro service in Changzhou now. However, over 260 bus lines have formed an effective bus system, and bus trips account for 12.2% of daily trips for all citizens in Changzhou<sup>[13]</sup>.

The smart card automatic fare collection (AFC) system was introduced in Changzhou in October 2001, which aims to decrease cash transactions. Passengers only need to tap their smart cards at the front door when boarding and all the buses have only one card reader at the front door. Besides, bus lines in Changzhou charge the same price, which is not distance-based. Either cash or smart cards can make the payment, but smart card users can enjoy a fare discount from 40% to 70%. Moreover, seniors over 70 years old can enjoy the bus services for free. It was reported that more than 75% of public transportation passengers in Changzhou used the smart card for payment in 2015. This proportion reached 95% in some of the bus lines. When a passenger is tapping on from a bus, the detailed transaction information is recorded, including time, bus ID, bus lines ID, smart card ID, the type of smart card and other operational data.

It should be noted that the transit system in Changzhou has an independent AFC system and the automatic vehicle location (AVL) system, which means that we cannot directly obtain the location of each boarding transaction. Hence, identifying the boarding location of each record is a significant issue for further studies in this research. Different from some other smart card systems in the world with complete information about boarding and alighting, the system in Changzhou, like most cities in China, can

only provide boarding records. Besides, there is no automatic passenger counter (APC) system in Changzhou, which means that we cannot obtain the number of onboard passengers for each moving bus. Note that passengers are restricted to boarding at the front door and alighting at the rear door.

### 1.2 Data collection

This study contains one-week anonymous smart card data from March, 16th to 22nd, 2015 in Changzhou. We collected more than  $4 \times 10^5$  smart card transaction records from six bus lines with a high proportion of elderly passengers and a high rate (around 90%) of those making payments by smart card. The reasons why we choose smart card data from six bus lines are as follows: Firstly, all these bus lines have a relatively high proportion of elderly passengers (over 20%). Secondly, a cash payment made by passengers can create a large time interval, in that case, we cannot apply the recorded time from consecutive smart card data to calculate the boarding gaps and use it to approximate passenger boarding time. Therefore, it is necessary to select data from a dataset which has few cash transactions.

It is acknowledged that there are more than five types of cards used in the transit system. However, students, the disabled and other types of passengers make up only an extremely small proportion of the total number of passengers, which are ignored in the analysis. This research only focuses on analysis and modeling of adult and elderly passengers.

Tab.1 shows the characteristics description of the operating buses. The contents include more than 20 fields of information, but only six fields are used in this study as listed in the table, including bus line ID, vehicle name, entrance type, length and number of doors. In Changzhou, most of the bus lines only have one type of vehicle.

**Tab.1** Bus characteristics description

Line ID	Vehicle name	Entrance type	Length/ m	Boarding door	Alighting door
14	Huanghai DD6120G13	1	12	1	1
12	Changlong YS6121QG	1	12	1	1
56	Jinlong KLQ6108GE3	2	10	1	1
16	Changlong YS6121NG	1	12	1	1
6	Changlong YS6900G	1	9	1	1
39	Jinlong KLQ6108GE3	2	10	1	1

Tab.2 presents the description of the smart card data structure. In our research, only four recorded fields are used to analyze passenger boarding behavior: Card type, tapping time, bus ID and bus line ID.

**Tab. 2** Smart card data structure description

Information	Description
Transaction ID	Records of each transaction
Smart card ID	Card code for each smart card
Card type	Records of the card user (adult, student, elderly, disabled, other)
Tapping time	Records of the time when one person is tapping on a card reader
Bus ID	Given the number of each operating bus
Bus line ID	Given the number of each bus route

Due to the requirement of identifying passenger boarding location, GPS data are also obtained from the AVL system. Tab. 3 shows the detailed contents of GPS data.

**Tab. 3** GPS data structure information

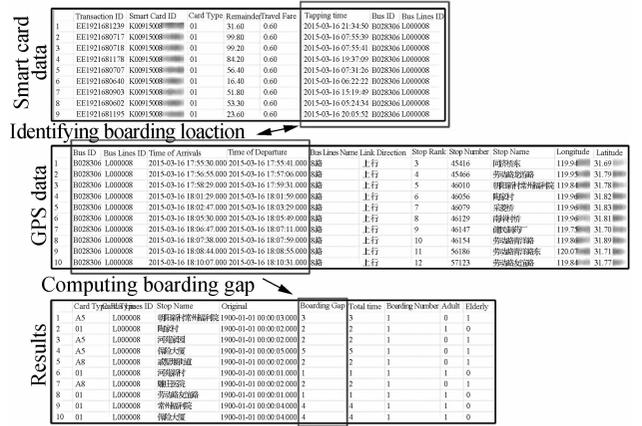
Information	Description
Bus ID	Given the number of each operating bus
Bus line ID	Given the number of each bus route
Time of arrival	Records of the bus arrival time at each stop
Time of departure	Records of the bus departure time at each stop
Longitude	Longitude records
Latitude	Latitude records

### 1.3 Data preprocessing

Data preprocessing is required before passenger boarding behavior analysis and modeling. The first step in data preprocessing is to determine the bus stop where the smart card users board. By linking the bus ID and bus line ID, we can narrow down the possibilities of the boarding location. Then, if we find that the tapping time in the smart card data is between the arrival time and departure time of one bus, we find the accurate boarding stop for this smart card record. The next step is to compute the boarding gap between two successive boarding records of two passengers. The computed boarding gap is the approximated boarding time for the latter passenger. In this way, we should notice that the boarding gap of the first passenger at each stop cannot be calculated, so the number of recorded activities is one fewer than the actual number of the boarding passengers. Since the smallest unit of the bus boarding gap is a second, the minimum gap is 1 s. Also, we will only use the terminology “boarding gap” to stand for the time that one passenger spends in boarding a bus in the rest of this research in order to avoid confusion. Fig. 1 shows a complete process of computing passenger gaps.

Some extreme large gaps are defined as outliers in this study. These outliers are mainly caused by the following factors: 1) Passengers making payments by cash (Approximately 10% of payments in the records are made by cash); 2) The passenger taking his card out after boarding; 3) Bus drivers waiting for the late-coming passengers. The clustering-based method is implemented to detect and clear the outliers.

Visual C# program language and structured query language (SQL) are applied to accomplish the processes above.

**Fig. 1** Process of computing boarding gaps

## 2 Analysis and Methodology

### 2.1 Descriptive analysis

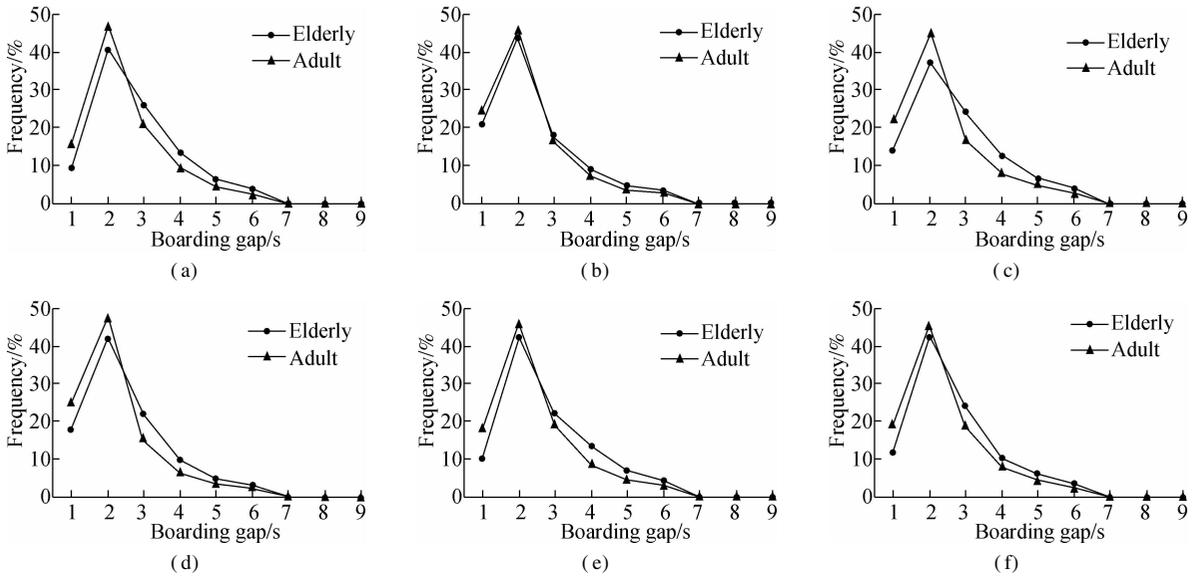
The critical issue of this study is to figure out if there is a statistically significant difference in the boarding gap between adult passengers and elderly passengers. Thus, we calculate the statistics of boarding gap for adult passengers and elderly passengers. The percentages of the boarding gap for both types of passengers are computed by dividing the total number of boarding records for one type of passenger by the number of different boarding gaps. The line charts in Fig. 2 show the percentages of boarding gaps for both adult and elderly passengers. The scale of the vertical axis reflects the value of the frequency distribution. The scale of the horizontal axis indicates time durations of different boarding gaps. As can be seen in Fig. 2, elderly passengers are likely to spend more time when boarding a bus.

Descriptive statistics (see Tab. 4) demonstrate that the average boarding gap of elderly passengers is higher than that of adult passengers. Then, we conduct a one-way analysis of variance (ANOVA) to further check the difference in their average boarding gaps by examining the variances of samples.

### 2.2 Analysis of variance

Analysis of variance is a hypothesis-testing technique used to analyze the differences among means. In this study, ANOVA allows us to determine whether there are systematic treatment effects that cause the mean in one group to differ from the mean in another. Note that using ANOVA to compare the means from two independent samples is equivalent to using a *t*-test to make this comparison. However, ANOVA seems to be more suitable for this research due to its conservativeness<sup>[14]</sup>. The results of ANOVA are presented in Tab. 5.

From these results, we can demonstrate that the difference in boarding gap between these two types of passengers reaches a significant level. Thus, we should take into account this difference in the following study.



**Fig. 2** Distribution of boarding gap. (a) Bus line 6 with one-step entrance; (b) Bus line 12 with one-step entrance; (c) Bus line 14 with one-step entrance; (d) Bus line 16 with one-step entrance; (e) Bus line 39 with two-step entrance; (f) Bus line 56 with two-step entrance

**Tab. 4** Statistics of passenger boarding gap

Passenger type	Number of records	Passenger boarding gap				
		Mean/s	Std. deviation/s	Std. error of the mean/s	Minimum/s	Maximum/s
Adult	152 537	2.46	1.323	0.009	1	6
Elderly	64 897	2.74	1.402	0.015	1	7
Total	217 434	2.54	1.353	0.008	1	7

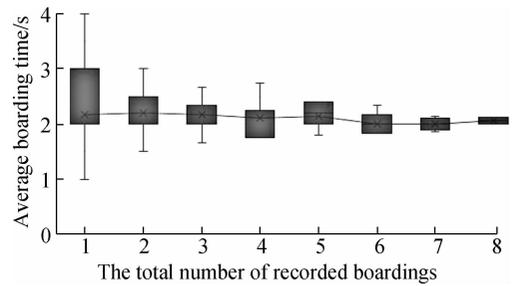
**Tab. 5** ANOVA for adult and elderly passengers

Type of comparison	Sum of squares	Degree of freedom	Mean square	F-statistic	p-value
Between groups	514.418	1	514.418	283.557	0
Within groups	394 435.230	217 432	1.814		
Total	394 949.648				

**2.3 Regression analysis**

Although the descriptive analysis in Section 2.1 has already revealed the variation of gaps to some extent, it is necessary to further explore changes in the dynamic boarding process before conducting the regression analysis. Fig. 3 presents the trends of the average boarding gap  $\bar{g}$  against the volume of boarding passengers at each bus stop. The average boarding gap  $\bar{g}$  has the highest variability with a large range of possible values for the average boarding gap when there is only one recorded boarding activity. The trend of this variation is clearly shown in the boxplot. As the number of passengers boarding a bus at a bus stop increases, the variability of the average boarding gap duration decreases and finally  $\bar{g}$  reaches a relatively stable value of 2 s. This pattern of variation is a reflection of the marginal effect of each extra boarding passenger on the dynamic process of boarding a bus. Considering the difference in boarding behaviors between adults and the elderly, we add the marginal effect for additional boarding of each type of passengers into our model. However, the marginal effects of adult and elderly passengers on  $\bar{g}$  can be reasonably complicated in this re-

search since adults and the elderly have mutual influences on each other when boarding buses. In other words, we can claim four or even more types of marginal effects when we study this process, namely, the marginal effects of each additional boarding on the average boarding gap of adult passengers and elderly passengers, respectively.



**Fig. 3** Trend of average boarding gap against the volume of boarding passengers

To simplify our analysis, we only consider two marginal effects: 1) The marginal effect of extra boarding (no matter what type of passengers they are) on the average boarding gap of adult passengers  $b_m^{adult}$ ; 2) The marginal effect of extra boarding on the average boarding gap of elderly passengers  $b_m^{elderly}$ .

Also, we include one dummy variable  $D$ , which is active to be one if one bus has a two-step entrance; otherwise  $D$  equals zero, representing a low floor entrance. One can expect that setting a two-step entrance in a bus increases passenger boarding gaps. Using a nonlinear regression model, we can measure the average cost of each

adult passenger boarding  $g_a$  and elderly passenger boarding  $g_e$ :

$$T = Ag_a + Eg_e + \varepsilon = A(g_b^{\text{adult}} + b_m^{\text{adult}}R + b_s^{\text{adult}}D) + E(g_b^{\text{elderly}} + b_m^{\text{elderly}}R + b_s^{\text{elderly}}D) + \varepsilon \quad (1)$$

where  $T = t_{\text{last}}^{\text{boarding}} - t_{\text{first}}^{\text{boarding}}$  denotes the total passenger boarding time at each bus stops;  $t_{\text{last}}^{\text{boarding}}$  and  $t_{\text{first}}^{\text{boarding}}$  are the time for the last passenger and the first passenger to board one bus, respectively;  $A$  is the number of recorded boarding activities of adult passengers for each stop;  $E$  represents the number of recorded activities of elderly passengers at each stop;  $R$  is the total number of recorded boarding passengers and  $R = A + E$ ;  $\varepsilon$  is the residual of the model;  $g_b^{\text{adult}}$  and  $g_b^{\text{elderly}}$  represent the basic average boarding gaps for adult passengers and elderly passengers, respectively;  $b_s^{\text{adult}}$  and  $b_s^{\text{elderly}}$  are the parameters indicating the contribution of bus entrance type to adult and elderly passengers. The regression results are shown in Tab. 6.

**Tab. 6** Results of regression analysis

Parameter	Estimate	<i>t</i> -test	<i>p</i> -value
$g_b^{\text{adult}} / (s \cdot \text{passenger}^{-1})$	2.210	63.040	0
$b_m^{\text{adult}} / (s \cdot \text{passenger}^{-2})$	-0.014	-7.542	0.001
$b_s^{\text{adult}} / (s \cdot \text{passenger}^{-1})$	0.074	6.332	
$g_b^{\text{elderly}} / (s \cdot \text{passenger}^{-1})$	2.700	61.808	
$b_m^{\text{elderly}} / (s \cdot \text{passenger}^{-2})$	-0.144	10.294	
$b_s^{\text{elderly}} / (s \cdot \text{passenger}^{-1})$	0.120	6.169	
Adjusted $R^2$		0.756	

From the modeling results, we find that there is a difference of 0.49 s/passenger between adults and seniors in the basic boarding gaps, namely  $g_b^{\text{adult}}$  and  $g_b^{\text{elderly}}$ , suggesting that even without considering the specific characteristics of the bus, seniors are slower to board buses. The effects due to the existence of a two-step entrance at the front door are significant, increasing both the average boarding gaps for adults by 0.074 s and seniors by 0.120 s, respectively. Note that there is a small distinction of 0.046 s between  $b_s^{\text{adult}}$  and  $b_s^{\text{elderly}}$ , suggesting that the effect of the step entrance is higher for the elderly. We also find that the total number of recorded boardings has a negative marginal impact on the average boarding gap for both adults and seniors, indicating that the average boarding gap decreases with the increase of boarding passengers, which is consistent with the findings obtained by Sun et al.<sup>[11]</sup>. There is a clear difference between  $b_m^{\text{adult}}$  and  $b_m^{\text{elderly}}$ , suggesting that the marginal effect is more significant for elderly passengers. To sum up, the process of tapping smart cards and waiting for reader responses accounts for a substantial part of the boarding time, and indeed seniors spend more time than adults in this process. Besides, the one-step entrance enables passengers to board more quickly than a two-step entrance by reducing the extra effort for passengers to walk up the steps, espe-

cially for the elderly.

### 3 Conclusion

In this paper, regression analysis is conducted using data from the automatic fare collection system in Changzhou, with the objective of quantifying the boarding gap and difference between adult and elderly passengers and analyzing the influence of the specific determinants of total boarding time at each bus stop, such as age and steps inside buses. The results of ANOVA demonstrate that there is a statistically significant difference between adults and seniors in the time spent on boarding a bus. In regression analysis, the model is estimated to account for the total time for a queue of passengers to boarding. We find that there are some negative marginal effects on both adults and seniors, indicating that a high demand for taking a bus speeds up the boarding process of a queue of passengers waiting at a bus stop by decreasing the average passenger boarding time. Moreover, this effect is greater on seniors than adults. From an operational aspect, this marginal effect brings extra variability in the boarding process and then this is expected to have an impact on bus service reliability. Besides, the existence of a two-step entrance is found to make the boarding process slower, which is an expected outcome. This negative impact is more significant in seniors than adults. Although the total boarding time cannot adequately represent bus dwelling time, it can be seen as an essential component of the time spent on transferring passengers. Some policy recommendations might be proposed based on the results of our study. Seniors usually spend more time boarding a bus and bring more variability to the dynamic boarding process. Thus, it might be a cost-effective way for bus agencies to provide customized bus services for the elderly, using low-floor buses with optimal capacity, to improve the efficiency of the whole transit system.

**Acknowledgments** The authors appreciate Changzhou Public Transportation Group Corporation for providing the data used in this study.

### References

- [1] Guenther R P, Hamat K. Transit dwell time under complex fare structure [J]. *Journal of Transportation Engineering*, 1988, **114** (3): 367 – 379. DOI: 10.1061/(asce)0733-947x(1988)114:3(367).
- [2] Levine J C, Torng G. Dwell-time effects of low-floor bus design [J]. *Journal of Transportation Engineering*, 1994, **120**(6): 914 – 929. DOI:10.1061/(asce)0733-947x(1994)120:6(914).
- [3] Dorbritz R, Lüthi M, Weidmann U, et al. Effects of on-board ticket sales on public transport reliability [J]. *Transportation Research Record; Journal of the Transportation Research Board*, 2009, **2110**: 112 – 119. DOI: 10.3141/2110-14.
- [4] Furth P G. *Data analysis for bus planning and monitoring*

- [M]. Washington, DC, USA: Transportation Research Board, 2000:5-15.
- [5] Tirachini A. Bus dwell time: The effect of different fare collection systems, bus floor level and age of passengers [J]. *Transportmetrica A: Transport Science*, 2013, **9** (1): 28-49. DOI:10.1080/18128602.2010.520277.
- [6] Bertini R L, El-Geneidy A M. Modeling transit trip time using archived bus dispatch system data [J]. *Journal of Transportation Engineering*, 2004, **130** (1): 56-67. DOI:10.1061/(asce)0733-947x(2004)130:1(56).
- [7] Shalaby A, Eng P, Farhan A, et al. Bus travel time prediction model for dynamic operations control and passenger information systems [C]//*The 82nd Annual Meeting of the Transportation Research Board*. Washington DC, USA, 2002.
- [8] Li M T, Zhao F, Chow L F, et al. Simulation model for estimating bus dwell time by simultaneously considering numbers of disembarking and boarding passengers [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2006, **1971**: 59-65. DOI: 10.3141/1971-09.
- [9] González E M, Romana M G, Álvaro O M. Bus dwell-time model of main urban route stops [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2012, **2274**: 126-134. DOI:10.3141/2274-14.
- [10] Widanapathirana R, Bunker J M, Bhaskar A. Modeling the BRT station capacity and queuing for all stopping busway operation [J]. *Public Transport*, 2015, **7**(1): 21-38. DOI:10.1007/s12469-014-0095-y.
- [11] Sun L J, Tirachini A, Axhausen K W, et al. Models of bus boarding and alighting dynamics [J]. *Transportation Research Part A: Policy and Practice*, 2014, **69**: 447-460. DOI:10.1016/j.tra.2014.09.007.
- [12] Zhao K J, Sun L J, Jin J G, et al. Analysis of crowding effect on passengers' movement time based on smart card data [C]//*15th COTA International Conference of Transportation Professionals*. Beijing, China, 2015: 1479-1489. DOI:10.1061/9780784479292.136.
- [13] Introduction of Changzhou Public Transportation Company [EB/OL]. (2009-04-05) [2016-12-18]. <http://www.czgj.cn/column1.aspx>.
- [14] Spiegelman C, Gates T. Post hoc quantile test for one-way analysis of variance using a double bootstrap method [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2005, **1908**: 19-25. DOI:10.3141/1908-03.

## 基于公交刷卡数据的老年乘客与普通成年乘客上车时间差别分析

雷 达 陈学武 程 龙 罗荣根

(东南大学交通学院, 南京 210096)

**摘要:**作为公交停车时间的重要组成,乘客上车时间对公交运行可靠性及公交服务水平具有重大影响.为研究老年乘客与成年乘客上车时间差异以及影响上车时间长短的因素,进而更好地理解乘客上车行为过程、减少乘客上车所花时间,提出了一种乘客上车时间回归分析模型.通过计算常州公交 IC 刷卡连续记录之间的时间差,得到了每位乘客上车时间的估计值,定义为上车时间间隔.然后,利用方差分析验证 2 种乘客的上车时间间隔是否存在显著差异,最后,建立回归模型分析乘客类型、每多一位乘客带来的边际效益及上车踏板数量对总上车时间的影响.研究表明,即使不考虑公交车型特征,普通成年乘客与老年乘客的上车时间间隔也存在一定差异;每一车次乘客平均上车时间随乘客数量的增加而减少;公交上车踏板数量越多,乘客尤其是老年乘客上车所花的时间越长.

**关键词:**老年人;公交刷卡数据;上车时间差别;方差分析;回归分析;边际效应

**中图分类号:**U121