

Estimating the number of taxi passengers depending on time zones and weather conditions

Dae Yong Lee^{1,1}, Min Kyu Park^{1,1}, Jong Seon Jang^{1,2}, Wan-Sup Cho^{2,3}

¹Dept. of Business Data Convergence

²Dept. of MIS/Business Data Convergence

Chungbuk National University, Chungbuk, Korea

¹{Ekddmf, ruserive}@gmail.com, ²asjoies@naver.com, ³wscho@chungbuk.ac.kr

Abstract. Although various analyses have been conducted for improving taxi imbalance of supply and demand, it is necessary to verify significant relation of variables deduced from analyzing process because existing researches were centered on apprehension of current condition. In this study, it was verified if time zones and weather conditions affected the number of taxi passengers on the basis of call taxi operating data for more than six months. Consequently, hypotheses were partially proved, and scientific taxi demand prediction is expected to be feasible based on the data analysis.

Keywords: Taxi demand, Time zones, Weather, ANOVA, Regression analysis

1 Introduction

As taxi has grown as a main method of transportation, growing problem of oversupply appeared because of taxi imbalance of supply and demand[1]. As a result, researches on optimal supply and polices appeared. However, Researches are not sufficient to analyze various variables affecting taxi demand. This study verified if time zone and weather conditions affected the number of taxi passengers based on call taxi operating data.

2 Related Work

One of their examples is a research which developed a model to calculate an optimal bus operation interval considering the effect of day of the week and time slot on bus passenger demand [2]. Another one is a research which verified the effect of weather, season, and holiday on the number of customers to contact on web-site [3].

As similar researches on the same field, there are one research which defined weather as its main weighting to calculate an optimal taxi wandering route [4] and the other one which analyzed bus operating pattern depending on weather change [5]. However, there was a limitation that direct effect of time slot and weather on taxi demand was not verified based on GPS spatial coordinate data.

3 Hypothesis and Experiment Design

We introduce hypothesis verification based on time zones and weather conditions.

3.1 Data Acquisition and Parameter Setting

This study made use of call taxi operating data[6] and regional detailed observation data of Meteorological Office from January to June in 2014. To collect taxi operation data, boarding places were extracted based on GPS spatial coordinate data of date and time whenever call taxi uses happened. In addition, to secure reliability, 546 lines of data were selected. In this study, Variables used are followings like Table 1 (time zone classifications) and weather conditions (rainfall, temperature, humidity etc.).

Table 1. 6 divisions of time zones.

Time division	Morn. Peak	Morn.	Aft.	Aft. Peak	Eve.	Night
Time reference	06:00 ~09:00	09:00 ~12:00	12:00 ~17:00	17:00 ~20:00	20:00 ~24:00	00:00 ~06:00

3.2 Hypothesis Formulation

Taxi demand can depending on various factors. We construct following hypotheses:

Hypothesis 1. $H_0 : T\mu_1 = T\mu_2 = T\mu_3 = T\mu_4 = T\mu_5 = T\mu_6$
 $T\mu_i$: average number of taxi passengers for the time zone i.

Hypothesis 2. $H_0 : R^2$ (Multiple R-squared) = 0
 R^2 : significance between weather conditions and taxi passengers.

3.3 Experiment Design

To verify hypothesis of Section 3.2, following experimental method was designed.

Experiment 1. ANOVA will be conducted to verify if there exist different numbers of taxi passengers depending on more than 3 time zones.

Experiment 2. Regression analysis will be conducted to verify correlation between the number of taxi passengers and weather condition (rainfall, temperature, humidity, wind speed).

4 Result of Analysis

4.1 Basic Statistics Analysis

Before verifying hypothesis, basic statistics used in this study was analyzed to figure out characteristics of the data used in analysis of actual proof. Table 2 shows average number of taxi passengers depending on the time zones. Table 3 shows basic statistics of the weather condition about rainfall, temperature, humidity, and wind speed when the taxi passenger data occurred in the past six month.

Table 2. Basic statistics of average number of taxi passengers by time zones.

	Morn. Peak	Morn.	Aft.	Aft. Peak	Eve.	Night
Mean	1.2593	1.9	2.2874	1.7826	1.7424	1.641
Std. Deviation	0.5257	1.4106	1.9404	1.0829	0.9815	1.0127
Variance	0.2764	1.9898	3.7653	1.1726	0.9634	1.0256
Sum	34	114	199	123	115	64
Observation	27	60	87	69	66	39

Table 3. Basic statistics of weather conditions.

	Rainfall	Temperature	Humidity	Wind Speed
Mean	0.1137	17.7767	55.9066	2.1454
Std. Deviation	0.7177	6.4489	21.9045	1.3823
Variance	0.515	41.5885	479.8059	1.9109
Observation	546	546	546	546

4.2 Actual Proof Analysis

First, Table 4 shows that null hypothesis (H_0) was rejected after verification of **Hypothesis 1** so $T\mu_1 \neq T\mu_2 \neq T\mu_3 \neq T\mu_4 \neq T\mu_5 \neq T\mu_6$ was verified.

Next, Table 6 shows that null hypothesis (H_0) was rejected after verification of **Hypothesis 2** so Multiple R-squared $\neq 0$ was verified. However, it is difficult to argue that H_1 as definitely proved with every factor of weather conditions because the $Pr(>|t|)$ values of rainfall, temperature and wind speed exceed the level of significance $\alpha = 0.05$. In addition, given data and regression straight line were correlated with 2.97%. In the other word, current data of its low cogency is not sufficient to explain the number of taxi passengers depending on weather conditions.

Table 4. Result of ANOVA

Level of significance $\alpha = 0.05$
Response: Number_of_Passengers

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Time_division	5	28.92	5.7833	3.1408	0.008707 **
Residuals	342	629.74	1.8413		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5. Result of polynomial regression analysis

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.552989	0.121827	12.748	< 2e-16 ***
Rainfall	0.041454	0.035103	1.181	0.238
Temperature	-0.005376	0.003894	-1.380	0.168
Humidity	-0.004818	0.001206	-3.994	7.38e-05 ***
Wind_speed	-0.001933	0.017914	-0.108	0.914

Signif. codes: 0 '***' 0.001, Multiple R-squared: 0.0297, p-value: 0.002595

5 Conclusion and Limitation

This study derives the following two conclusion: Firstly, ANOVA analysis verified that the number of taxi passengers showed difference depending on time zones. Secondly, it was verified that weather conditions were not sufficient to explain the number of taxi passengers through polynomial regression analysis.

Above result has a significance that it was based on the real traffic data collected from on-line taxi company. It has a limitation on the side of representativeness that small amount of data was used for verification. If sufficient amount of data can be used for analysis, more precise verification result will be deducted.

Acknowledgments. This research was supported by the MSIP(The Ministry of Science,ICT and Future Planning), Korea, under the "SW master's course of a hiring contract" support program (NIPA-2014-HB301-14-1011) and ITRC(Information Technology Research Center) support program (NIPA-2014-H0301-14-1022) supervised by the NIPA(National IT Industry Promotion Agency)

References

1. S.W Kang, Road traffic: administration business, Public Officials Benefit Association, pp.64-67, vol.44 No.482(2009)
2. S.Y Kho, J.S Kho, A Development of Optimal Bus Headway and Fleet Size Model, Journal of Korean Society of Transportation, vol.16 No.2(1998)
3. D.H Kim, A Study on the Influence of Weather and Weekly Cycle Characteristics on the Provision of Tourism Information, Major of MIS, Hanyang University (2008)
4. J.H Jeong, Optimal roaming route recommendation by analyzing operation record of taxi, Department of Business Data Convergence, Chungbuk National University(2014)
5. S.Y Nam, B.S Son, A Study on Bus driving Pattern due to Weather changes, Proceedings of the KOR-KST Conference, vol.2012 No.1(2012)
6. Dangol Taxi, <http://dangoltaxi.com/>