Estimating Pedestrian Traffic Volume: A Preliminary Analysis

Dae-Jin KIM [a], Joonho KO [b], Young-ihn LEE [c]

[a,b] *Department of Transportation System Research, The Seoul Institute, Seoul, 137-071, Korea*
[a] *E-mail: syncmkdj@si.re.kr*
[b] *E-mail: jko@si.re.kr*
[c] *Graduate School of Environmental Studies, Seoul National University, Seoul, 151-742, Korea*
[c] *E-mail: yilee@snu.ac.kr*

Abstract: In recent years, pedestrian environments have become more important in transportation policies. To form a policy for pedestrians a survey on pedestrian volume needs to be fulfilled in advance, however it spends a lot of cost and time. Therefore, this study aims to conduct a preliminary analysis to estimate pedestrian traffic volume. In this study, a lot of variables were considered such as population, employees, public transits and floor area. From some official institutes, the data for variables were obtained. The pedestrian flow rates by time of day show that there are three peak times: a.m. peak, noon peak and p.m. peak. From the results of correlation analysis, the total number of employees and accessibility to public transits seem to be closely associated with pedestrian traffic volumes. This study provides a framework for estimating pedestrian traffic volume.

Keywords: Pedestrian, Volume, Estimate, Correlation, Seoul

1. INTRODUCTION

In recent years, pedestrian environments have become a major part of transportation policies with increasing concerns about global warming. Walking is eco-friendly travel mode, which does not emit any gases that can harm the environment. In Korea, however, a government has been developed mostly focused on economic growth; it has led to increasing number of cars on roads during past 50 years. Even though principal travel modes are pedestrian walking, the policies in the fields of transportation have primarily focused on cars. Consequently, the walking environments have become worse than the other transportation environments. For this reason, Korean researchers in the fields of transportation started studying on non-motorized travel modes, especially pedestrian movements. Those had tried to analyze pedestrian safety, satisfaction and movements. Most of those studies noted that a survey on the pedestrian volume needs to be fulfilled in advance. However, a survey on counting the pedestrian volume spends a lot of cost and time. In order to overcome those restrictions, some recent studies have made an effort to establish a pedestrian volume model in order to save cost and time, but it has not been effectively done. Therefore this study's purpose was to find a good model to estimate pedestrian traffic volume, and to establish a data set that is supposed to be used in the model.

Considering the current status of database construction in Korea, the most suitable model could be sketch-plan model. The sketch-plan models in many previous studies have predicted the pedestrian volume through the use of pedestrian counts and regression analysis as a function of adjacent land uses and/or indicators of transportation trip generation. It has been known that this model is cost-effective and could be applied in a variety range of space units.

The objective of this study is to fulfill a preliminary research on estimating pedestrian traffic volumes. Thus, this study tried to find good variables that explain the pedestrian traffic volumes well in advance. The data of explanatory variables were collected and processed from some public research institutes in Korea. Then, some basic statistical analyses were fulfilled to know about the relationship between the pedestrian volume and explanatory variables.

2. LITERATURE REVIEW

A lot of methods have been used to estimate pedestrian volume. There are two models that use regression modeling techniques: sketch-plan and space-syntax. The sketch-plan model predicts pedestrian volumes through the use of pedestrian counts and regression analysis as a function of adjacent land uses and/or indicators of transportation trip generation (Clifton et al., 2004). Several studies defined some factors that are associated with pedestrian volume as shown in Table 1. Schneider et al. (2008) created a pilot model of pedestrian intersection crossing volumes in Alameda County, and the model of this study showed a good overall fit (adjusted-R^2=0.897). This model consisted of surrounding land uses, transportation system attributes, and neighborhood socioeconomic characteristics. Miranda-Moreno and Fernandes (2011) investigated a relationship between the natural log of pedestrian activity and some explanatory variables such as land use, transit network patterns and weather conditions. The models in this study also showed good explanation abilities by using log-linear forms (the ranges of adjusted-R^2 were from 0.53 to 0.60). Kim and Kim (2011) fulfilled a study on determining factors on pedestrian volume by station area types by using regression analysis. The models used in this study seemed to have good model fits with high F-values. However, the models seemed not to have good explanation power to estimate pedestrian volumes: the average value of R^2 was 0.153. The advantage of sketch-plan model is that this model can estimate the pedestrian volumes by a small cost. In addition, it can be applied in a variety range of space units in urban areas. Even though the sketch-plan models in previous studies showed good results, however, it is generally believed that the sketch-plan model is hard to obtain a statistical confidence and also difficult to assign the estimated volumes to specific streets.

Table 1. Factors associated with pedestrian volume

Classification	Variables
Land use	Population, Employment, Company, Building area, Type of land
Transportation system	Proximity to subway, Proximity to bus stop, Alighting & Boarding passengers nearby subway, Alighting & Boarding passengers nearby bus stop, The number of bus stop nearby, The number of subway nearby, Nearby road length, The number of traffic land nearby, Proximity to crosswalk,
Pedestrian street feature	Width, Slope, Obstacle, Material, Facility for disabled person

Source: Schneider et al. (2008)

Space-syntax model selects the space-syntax network analysis modeling, which calculates the accessibility from a whole space to one specific space; this is called integration. The space-syntax model also uses regression modeling to estimate the pedestrian volumes. Several studies have proved that the space-syntax models predicted pedestrian volumes with a high level of accuracy (Radford and Ragland, 2004; Desyllas et al., 2003). However, the space-syntax model requires special software, such as Axman (3.0) or Fathom Visibility Graph Analysis Software, which restricts researchers from using their own abilities.

Finally, route choice model has also been used in a lot of studies. This model followed the conventional four-step travel model that has been widely used in the transportation planning field. This model is composed of four steps: trip generation, trip distribution, modal split, traffic assignment. For a long time, the four-step model has been commonly used when estimating the motorized vehicles. However, some studies emphasized that several other steps should be taken into account when estimating a non-motorized transportation mode (Schneider et al., 2011; Jones and Buckland, 2007). In addition, this model was typically constructed with traffic analysis zones that are too large to reflect pedestrian activities. It seems that this model would not be appropriately applied in Korea because most Korean transportation database not only has been based on large scale zone systems, but also has not had a detailed network database of pedestrian streets.

The literature indicates that the pedestrian volume models could be classified into two models: regression and route choice modeling. Each model has advantages and disadvantages considering cost-effectiveness and explanation power. Regression models are cost-effective, but are not typically calibrated to actual pedestrian volume counts. However, route choice needs a lot of efforts to obtain database for estimating volumes, but it could have more accurate results and wide applications. In Korea, though, the regression modeling could be preferred than route choice considering the current state of data basement of pedestrian activities. In fact, present data of pedestrian trips have not been established in detail in Korea, yet.

3. DATA

3.1 Description of variables

In order to build a model to estimate the pedestrian traffic volume, this study considered a lot of variables that are associated with pedestrian volume as shown in Table 2. The data of variables were obtained from some public research institutes. Some data were imported into a Geographical Information System (GIS) and processed.

Table 2. Description of variables

Classification	Variables name		Description
Dependent variables			
Pedestrian volume	VOL	DAYVOL	all day pedestrian volume (7:00~21:00)
		AMVOL	am peak pedestrian volume (7:45～9:45)
		NNVOL	noon peak pedestrian volume (12:00～14:00)
		PMVOL	pm peak pedestrian volume (17:45～19:45)
Explanatory variables			
Land use	TOTPOP		total populations within a radius of 300 meters
	TOTEMP		total number of employees within a radius of 300 meters
	Floor area within a radius of 150m	ARESIDENTIAL	total floor area of residential buildings (km^2)
		ACOMMERCIAL	total floor area of commercial buildings (km^2)
		ABUSINESS	total floor area of business buildings (km^2)
		AMULTI	total floor area of multi-use buildings (km^2)
		AOTHER	total floor area of the other buildings (km^2)
	Land use dummy	RESIDENTIAL_DUM	residential district dummy (1=yes; 0=no)
		COMMERCIAL_DUM	commercial district dummy (1=yes; 0=no)
		GREEN_DUM	green district dummy (1=yes; 0=no)
		MANUFAC_DUM	manufacturing district dummy (1=yes; 0=no)
		CBD_DUM	central business district dummy (1=yes; 0=no)
Transportation system	SUBGRAV		accessibility to subway by gravity index
	BUSGRAV		accessibility to bus by gravity index
	NSUBWAY		the number of subway stations within a radius of 500 meters
	NBUSSTOP		the number of bus-stops within a radius of 300 meters
Pedestrian street feature	WIDTH		width of pedestrian street
	NVEHLANE		the number of lanes nearby vehicle road
	SLOPE_DUM		sloping pedestrian street dummy (1=yes, 0=no)
	BUS_DUM		bus stop dummy within 50 meters (1=yes, 0=no)
	SUB_DUM		subway station entrance dummy within 50 meters (1=yes, 0=no)
	CRO_DUM		crosswalk dummy within 50 meters (1=yes, 0=no)
	Street type	PEDONLY_DUM	pedestrian only street dummy (1=yes, 0=no)
		BICYCLE_DUM	bicycle and pedestrian street dummy (1=yes, 0=no)
		MIXTRAFFIC_DUM	mixed traffic street dummy (1=yes, 0=no)

3.2 Data collection

Pedestrian volume This study used pedestrian volume data collected by the Seoul Metropolitan Government in 2009 over 9,850 locations within the city of Seoul. The surveyed spots were evenly selected over Seoul all around as shown in Figure 1. This volume data were surveyed on pedestrian streets from 7 a.m. to 21 p.m. The surveyed spots were selected on the pedestrian street along main arterial roads over a width of 12 meters.

During the investigation period, the surveyors counted the number of pedestrians for 5 minutes, and then took a break for 10 minutes. In order to provide continuing volume data, the number of counts for 5 minutes was multiplied by three.

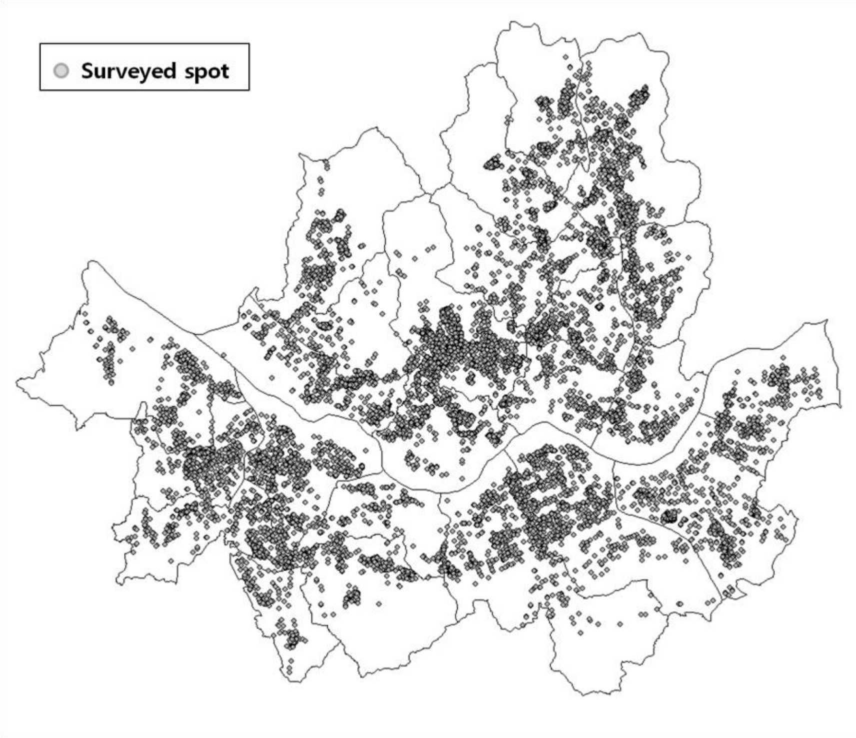

Figure 1. Surveyed spots of pedestrian volume

Population, number of employees and floor area The population and the number of employee data were obtained from a national statistics office. This data have been published every 5 years (the latest data were published at 2010) since 2000. The data were aggregated based on a special census track, and Seoul has 16,479 census tracks as represented Figure 2 (average area of each census track is 0.037 km^2 ; total area of Seoul is 605 km^2).

Figure 2. Census tracks (16,479 units)

This study obtained the population and the number of employees within a radius of 300 meters. For this, the census tracks that are within a radius of 300 meters were collected by using ArcGIS 9.2 tool. Then population and the number of employees in each census track were summarized. The population and the number of employees in particular census tracks that lie on the radius curve were calculated in proportion to the percentage of the area size inside the radius curve to the original area size of the census track. The equation of this process is represented as:

$$TOTPOP_r = \sum_s \frac{AREA_{s,r}}{AREA_s} \times POP_s \quad (1)$$

Where, $TOTPOP_r$: Population (or the number of employees) inside a buffer zone r

POP_s : Population (or the number of employees) in a particular census track s

$AREA_{s,r}$: Area of a particular census track s inside a buffer zone r

$AREA_s$: Original area of a particular census track s

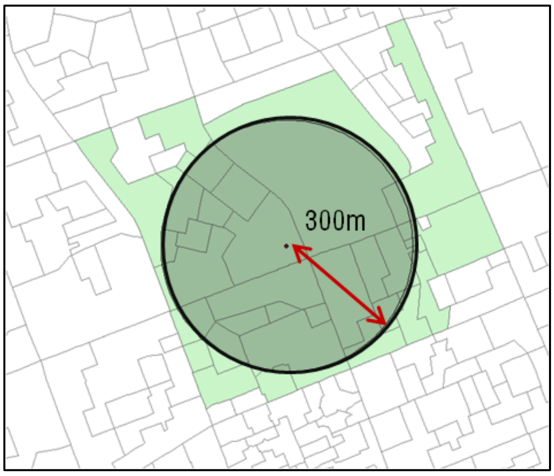

Figure 3. Example of selecting census tracks and calculating population inside a buffer zone

The data of floor area of buildings are collected based on an edited cadastral map. It has too various building types (45 building types), so it needs to be classified into some small groups. This study classified the building types into five groups: residential, commercial, business, multi-use and other buildings. The method to obtain the floor area within 150m buffer zone is similar to calculating the population. By using ArcGIS 9.2 tool, this study calculated the each type's floor area, respectively.

Land uses The information of land use districts are offered based on GIS information as shown in Figure 4. The type of land use districts can be classified into five groups at macro levels: residential, commercial, manufacturing, green, CBD (central business district). The type of land use district of each surveyed spot was then determined by intersecting with a map of the land use type.

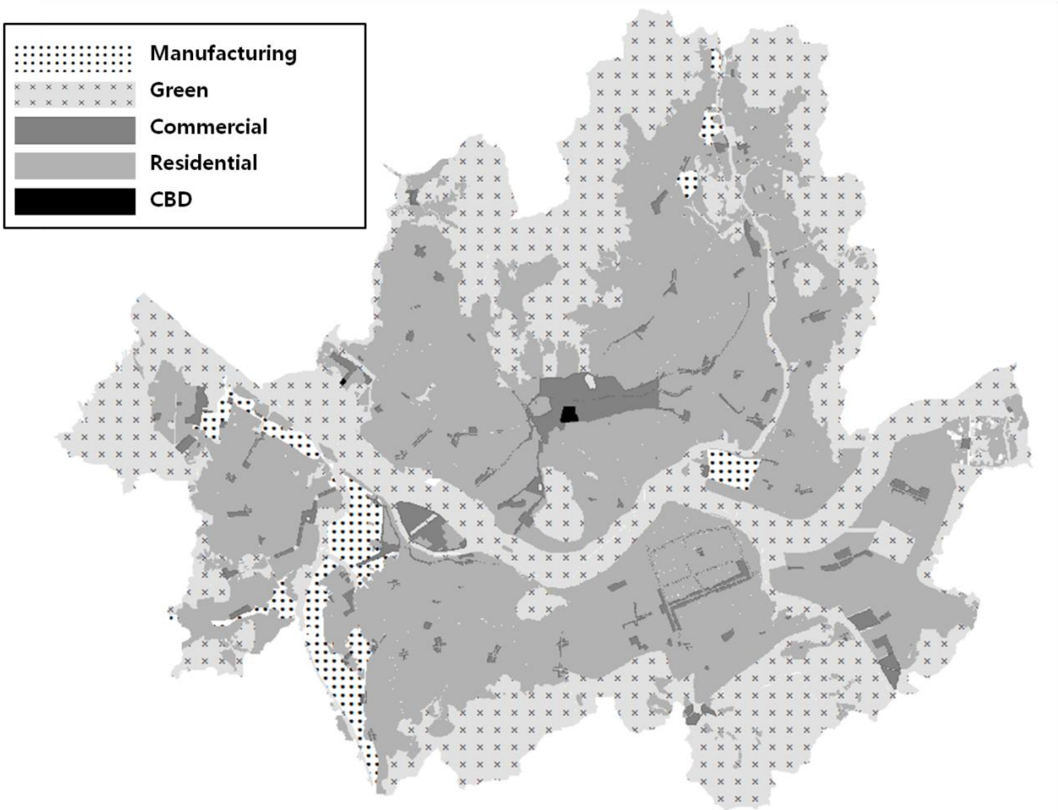

Figure 4. Land use districts in Seoul

Subway station and Bus stop Information about a subway and a bus were imported into the ArcGIS geographic information system viewer as shown in Figure 5 and Figure 6. The radius of catchment areas for subway stations and bus stops were defined 500 meters and 300 meters, respectively. The catchment area for subway station is especially perceived as 500 meters; though, the distances between bus stops are shorter than the one between subway stations and also the catchment area for bus stops is smaller than the one for subway stations. That was why this study chosen the different catchment areas of a subway station and a bus stop. The numbers of subway stations and bus stops within each buffer zone were calculated through the ArcGIS 9.2 tool.

An accessibility index to the subway and bus was established by a gravity index that was used in some previous studies (Brandt and Meannig, 2011; Bowes and Ihlanfeldt, 2001). In order to establish the gravity index, this study used the total number of boarding and alighting passenger data (obtained from transportation card data) at subway stations and bus stops from July to November at 2009. Then, the distances between the surveyed spots and subway stations (or bus stops) within a buffer zone were calculated by using ArcGIS 9.2 tool.

$$SUBGRAV_a = \sum_a \frac{G_b}{dist_{a,b}} \tag{2}$$

Where, $SUBGRAV_a$: Gravity index that indicates the accessibility to subway (or bus)

G_b : Total number of boarding and alighting passengers at subway station b (or bus stop) inside buffer zone

$dist_{a,b}$: Euclidean distance between the surveyed spot a and subway station b (or bus stop)

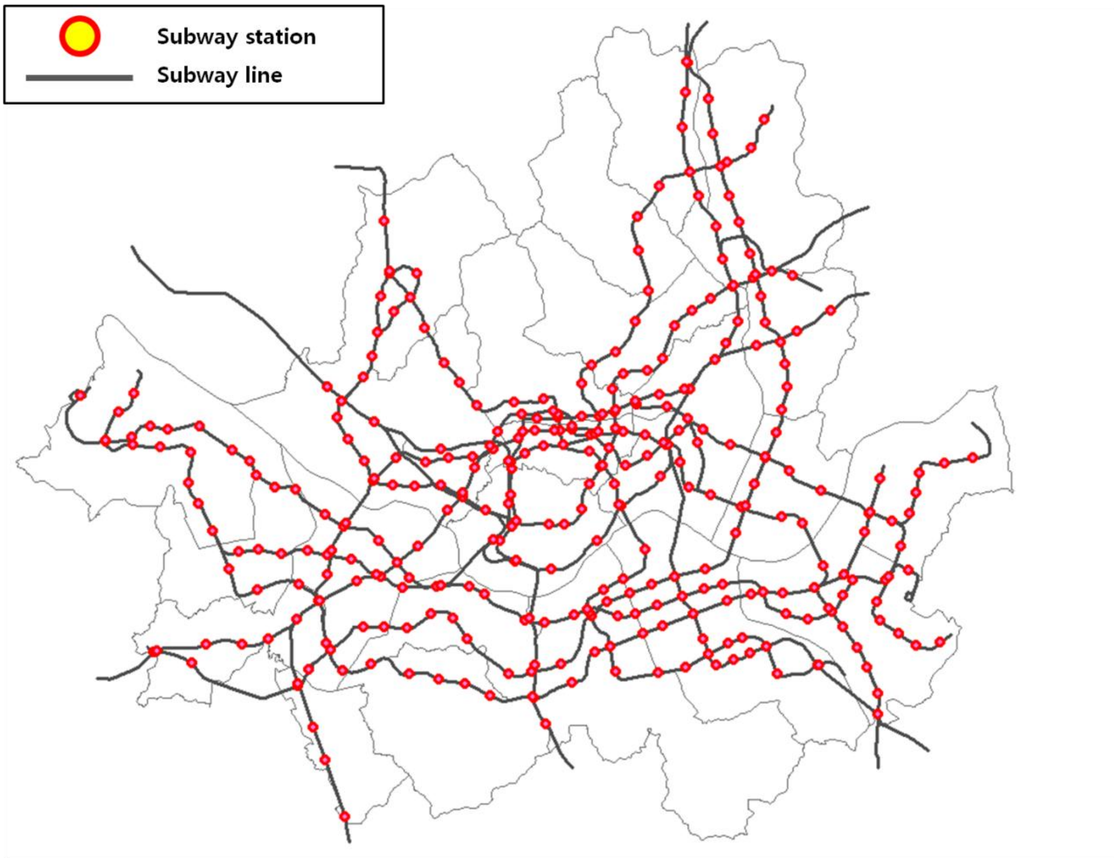

Figure 5. Subway stations and Subway lines in Seoul

Figure 6. Bus stops in Seoul

4. ANALYSIS

4.1 Pedestrian volume by time of day

The average pedestrian flow rates for 15 minutes of total 9,850 surveyed spots were represented in following Figure 7. It shows that the pedestrian volume is gradually increasing from morning to evening and also has three peak times in the morning, noon and evening respectively.

This study calculated 2-hour peak time by selecting the maximum level of pedestrian flow rate for eight consecutive 15-minutes. The result of the a.m. peak time was from 7:45 to 9:45, and next the noon peak time was from 12:00 to 14:00, and finally the p.m. peak time was from 17:45 to 19:45. It seems that changes of trip purpose along with the flow of time are reflected to the pedestrian flow rates. In Korea, most of workers and students attend the office and school before 9 a.m. in the morning. The graph of the pedestrian flow rates either shows that most of pedestrians were concentrated for one hour before 9 a.m. The noon peak time seems to exist when the workers travel along the way to have lunch and some business trips are produced through walking. The evening time pedestrian volume seemed to be higher than in other time periods. Especially, the p.m. peak time (from 17:45 to 19:45) is the time that leaving the office trips begin and leisure activities are being produced (such as shopping). The result of pedestrian flow rate represents that the pedestrian estimating model needs to be built in three types: am peak, noon peak and pm peak models.

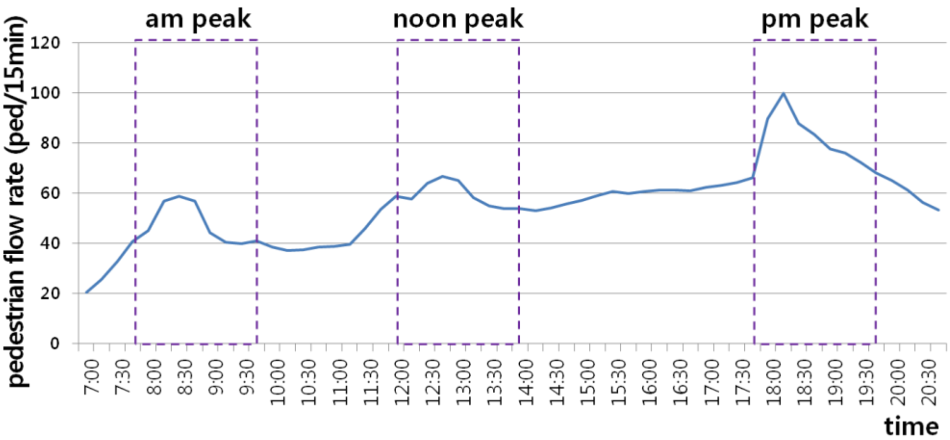

Figure 7. Average pedestrian flow rates

4.2 Descriptive statistics

Table 3 shows variable descriptive statistics of each variable. The average pedestrian volume during all-day was 2,986 pedestrians/14hr, and the range of it was from 7 to 113,606 pedestrians/14hour. The average volume of p.m. peak was bigger than the volumes during other times (a.m. peak, noon peak). Population and the number of employees varied according to the survey spot. The variations of those variables were quite big (the minimum of population and the number of employees were 21 and 20, respectively; the maximum of those were 28,561 and 99,351, respectively). Moreover, the variation of the number of employees was bigger than that of population. The average width of streets was about 4 meters, which is sufficiently wide for pedestrians to walk on; however, the widths of streets in some places were not appropriate to travel (the minimum width was only 1 meter).

Table 3. Variable descriptive statistics

	minimum	maximum	average	standard deviation
DAYVOL	7.00	113606.00	2986.64	3786.42
AMVOL	0.00	7042.00	397.95	477.04
NNVOL	0.00	11328.00	467.63	613.68
PMVOL	0.00	34702.00	623.06	887.69
TOTPOP	21.00	28561.00	9889.69	4792.91
TOTEMP	20.00	99351.00	9689.33	11253.90
ARESIDENTIAL	0.00	153.89	3.75	8.01
ACOMMERCIAL	0.00	80.46	1.51	2.99
ABUSINESS	0.00	28.78	0.79	1.39
AMULTI	0.00	30.38	0.52	1.39
AOTHER	0.00	40.90	0.49	1.58
RESIDENTIAL_DUM	0.00	1.00	0.77	0.42
COMMERCIAL_DUM	0.00	1.00	0.15	0.35
GREEN_DUM	0.00	1.00	0.03	0.17
MANUFAC_DUM	0.00	1.00	0.05	0.23
CBD_DUM	0.00	1.00	0.00	0.05
NSUBWAY	0.00	4.00	0.77	0.67
NBUSSTOP	0.00	20.00	5.72	3.29
WIDTH	1.00	24.30	3.96	2.16
NVEHLANE	1.00	18.00	2.88	2.40
SLOPE_DUM	0.00	1.00	0.25	0.43
BUS_DUM	0.00	1.00	0.23	0.42
SUB_DUM	0.00	1.00	0.07	0.25
CRO_DUM	0.00	1.00	0.46	0.50
PEDONLY_DUM	0.00	1.00	0.55	0.50
BICYCLE_DUM	0.00	1.00	0.05	0.22
MIXTRAFFIC_DUM	0.00	1.00	0.40	0.49

- Total number of samples : 9,850

4.3 Correlation analysis

From the results of correlation coefficients, we could know some variables that affect on pedestrian volumes. The total number of employees seemed to affect more on the pedestrian volume than the total number of populations. In addition, employees seemed to be closely related with pedestrian volumes at noon peak time (Pearson coefficient = 0.518), it may because pedestrian movements of employees for dinner have influence on pedestrian volumes. In correlation results, total population seemed to make negative effects on the pedestrian volumes. However, in this case this effect should not be considered that increasing population decreases the number of pedestrians; this effect could be adjusted by other variables by using advanced analysis methods. In this step, we can only assume that the effect of population is not closely related with the pedestrian volume in comparison with the number of employees. Accessibility to public transits also seems to be closely associated with pedestrian traffic volume. In land use district dummy variables, commercial and central business district seemed to have little positive relationships with the pedestrian volume. The number of subway and existence of subway station entrance also seemed to have little positive effects on pedestrian volumes. The subway facilities seem to have little correlation with pedestrian volumes. Among street type variables, the pedestrian only street has little positive effect on volume, but the other types of streets have negative effects (bicycle & pedestrian, mixed traffic street).

Table 4. Results of correlation coefficients

	DAYVOL	AMVOL	NNVOL	PMVOL
TOTPOP	-0.145**	-0.098**	-0.285**	-0.129**
TOTEMP	0.339**	0.306**	0.518**	0.316**
ARESIDENTIAL	-0.075**	-0.066**	-0.101**	-0.063**
ACOMMERCIAL	0.075**	0.036**	0.079**	0.069**
ABUSINESS	0.123**	0.123**	0.234**	0.118**
AMULTI	-0.017	-0.025*	-0.040**	-0.009
AOTHER	0.023*	0.057**	0.058**	0.022*
RESIDENTIAL_DUM	-0.191**	-0.166**	-0.288**	-0.164**
COMMERCIAL_DUM	0.242**	0.190**	0.362**	0.208**
GREEN_DUM	-0.082**	-0.085**	-0.088**	-0.077**
MANUFAC_DUM	-0.027**	0.054**	-0.027**	-0.025*
CBD_DUM	0.292**	0.098**	0.275**	0.277**
SUBGRAV	0.433**	0.397**	0.512**	0.497**
BUSGRAV	0.228**	0.400**	0.443**	0.512**
NSUBWAY	0.262**	0.211**	0.325**	0.239**
NBUSSTOP	0.152**	0.104**	0.158**	0.138**
WIDTH	0.167**	0.147**	0.164**	0.159**
NVEHLANE	0.204**	0.255**	0.149**	0.187**
SLOPE_DUM	-0.074**	-0.044**	-0.049**	-0.072**
BUS_DUM	0.141**	0.180**	0.098**	0.124**
SUB_DUM	0.259**	0.313**	0.212**	0.245**
CRO_DUM	0.099**	0.134**	0.082**	0.084**
PEDONLY_DUM	0.139**	0.158**	0.130**	0.116**
BICYCLE_DUM	-0.021*	-0.008	-0.045**	-0.019
MIXTRAFFIC_DUM	-0.131**	-0.156**	-0.112**	-0.109**

- ** : significant at $\rho < 0.01$, * : significant at $\rho < 0.05$

5. Conclusion

5.1 Findings

This study aimed to fulfill a preliminary research on estimating pedestrian traffic volume. In this study some basic data were obtained from research institutes. From those data, we could know the pattern of pedestrian traffic volume by time of day, and the relationships between pedestrian traffic volumes and explanatory variables. Among several variables, the total number of employees and accessibility to public transits seem to be closely associated with pedestrian traffic volumes. Increasing numbers of subway facilities, especially, seems to bring much pedestrian to the streets. The type of land use district is also related with pedestrian traffic volumes; the commercial and central business districts have little positive effects, but the other districts seem to have negative effects. Among street type variables, the pedestrian only street has positive effect on volume, but the other types of streets seem to have negative effects (bicycle & pedestrian, mixed traffic street).

In conclusion, this study seems to show the effect of each variable to pedestrian volumes. This study seems to be better than previous studies due to its rich dataset. In addition, the methods for building data set of this study could contribute to further studies when estimating pedestrian volumes in other cities.

5.2 Future studies

In future study, we could build a pedestrian estimating model by using data collected in this study. In order to build the model, some need to be considered.

First, the model forms should be determined in advance. There were several types of model forms using regression modeling in previous studies: linear, log-linear and the other possible forms. In general, the log-linear form has been widely used in many studies in order to avoid predicting negative values. However, the model form, which reflects the data of this study well, should be determined with careful consideration.

Second, the model could be divided by the type of land use district. From the results of correlation coefficients, we could know that the type of land use district has an effect on the pedestrian traffic volumes. The model could be divided into five types: residential, commercial, green, manufacturing and central business district models.

ACKNOWLEDGEMENTS

The authors are grateful to the Ministry of Land Infrastructure and Transport for their financial support (the project number is 12 high-tech city C18).

REFERENCES

Bowes, D. and Ihlanfeldt, K. (2001) Identifying the impacts of rail transit stations on residential property values, *Journal of Urban Economics*, Volume 50, Issue 1, 1-25.

Brandt, S. and Maennig, W. (2011) The impact of rail access on condominium prices in Hamburg, *Transportation*, Volume 39, Issue 5, 997-1017.

Clifton, K., Davies, G., Allen, W. and Radford, N. (2004) *Pedestrian Flow Modeling for Prototypical Maryland Cities*, Maryland Department of Transportation, Division of Highway Safety Programs, Hanover, MD.

Desyllas, J., Duxbury, E., Ward, J. and Smith, A. (2003) *Pedestrian Demand Modeling of Large Cities: An Applied Example from London*, Working paper 62, Center for Advanced Spatial Analysis, University College London.

Jones, M. and Buckland, L. (2007) Estimating Bicycle and Pedestrian Demand in San Diego. *Journal of the Transportation Research Board*, Washington DC. Available online: http://safetrec.berkeley.edu/news/08-2609.pdf

Kim, T., Kim, J. (2011) *A Study on Determining Factors on Pedestrian Volume by Station Area Types*. Seoul Development Institute, Seoul. Available online: http://www.sdi.re.kr/rept/resh/resh_viw.jsp?reportId=22294

Miranda-Moreno, L. F. and Fernandes, D. (2011) Modeling of Pedestrian Activity at

Signalized Intersections: Land Use, Urban Form, Weather, and Spatiotemporal Patterns. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 2264, 74-82.

Raford, N. and Ragland, D. (2004) Space Syntax: Innovative Pedestrian Volume Modeling Tool for Pedestrian Safety. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1878, Washington D.C., 66-74.

Schneider R. J., Arnold, L. S. and Ragland, D. R. (2008) A Pilot Model for Estimating Pedestrian Intersection Crossing Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 2140, 13-26.

Schneider, R. J., Henry, T., Mitman, M. F., Stonehill, L., Koehler, J. (2011) Development and Application of the San Francisco Pedestrian Intersection Volume, *Transportation Research Record: Journal of the Transportation Research Board*, Volume 2299, 2012, 65–78.