Investigating Pedestrian Walkability using a Multitude of Seoul Data Sources

Won Do Lee[a,b], Wim Ectors[c], Tom Bellemans[c], Bruno Kochan[c], Davy Janssens[c], Geert Wets[c], Keechoo Choi[a] and Chang-Hyeon Joh[d*]

a Department of Transportation Engineering, Ajou University, 16499 Suwon, Republic of Korea

b Crime and Well-being Big Data Centre, Manchester Metropolitan University, M15 6BH Manchester, England.

c Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium.

d Department of Geography, Kyung Hee University, Kyungheedae-ro 26, Dongdaemun-gu, 02447 Seoul, Republic of Korea

*Corresponding author: Chang-Hyeon Joh (Tel: +82-2-961-9264, Email: bwchjoh@khu.ac.kr).

Investigating Pedestrian Walkability using a Multitude of Seoul Data Sources

Currently walking is a multidisciplinary and emerging point of attention for urban sustainability and for ensuring the quality of pedestrian environments. In order to understand pedestrian behaviour, walkability researches estimate the factors which affect the level of pedestrian satisfaction. Past studies focused on the relationship between environmental factors and pedestrian behavioural outcomes. In this study, we developed pedestrian satisfaction multinomial logit models using various datasets, examining the relative impact of five differently themed sets of attributes: personal, walk-facilities, land-use, pedestrian volumes, and weather-related variables. The results show that the personal variability attributes were selected as most significant. We investigated effects of personal variability, such as the spatial cognition level and travel purpose, and detailed effects of environmental features. In addition, crowdedness, land-use types, and residential information were investigated. The results from this study offer contributions by providing evidence of the importance of personal and contextual variables in influencing the pedestrian walkability.

Introduction

Walking is the most fundamental transport mode. A suitable and qualitative walking environment is a vital part of a sustainable city (Lindelöw et al. 2014). Recently, research on walking has extended into a truly multidisciplinary research.

Especially in transportation studies, understanding walking behaviour is an emerging issue because walking is currently one of the primary and active transport modes, and an important consideration for sustainable cities (Antonini et al. 2006). Walking behaviour has increasingly been receiving attention in recent studies since walking is not only decreasing the car dependency (Gilderbloom et al. 2015), but also yields advantages towards keeping the city sustainable and enhancing the urban inhabitants' physical health (Carlson et al. 2015; Kang, 2015). Several recent studies employ simulation approaches and new technologies to understand pedestrian walking behaviour and to capture the pedestrian streams (Gehrke & Clifton 2017; Hussein & Sayed 2017; Yin & Wang 2016; Xie & Wong 2015; Nurul Habib et al. 2014).

Current walking behaviour researches focus on the "Walkability" which estimates the relationship between environmental factors and behavioural outcomes from pedestrians. Walkability is not only crucial to understand the pedestrian walking behaviour, but also became a significant indicator for estimating the level of pedestrian satisfaction. In addition, it may also be used to benchmark pedestrian facility quality by means of the pedestrian level of service (LOS).

The pedestrian satisfaction has been studied as an important indicator and various definitions have been given in empirical studies. Kim et al. (2014) reported that the pedestrian's psychological state might depend on individual experience, such as walking-friendly facilities and preceding travel experience. The conceptual idea of

pedestrian walkability in this study is likely to match the concept of pedestrian satisfaction, which is associated with personal and environmental factors.

There are two primary approaches in previous walkability researches to identify the significant factors which influence pedestrian walkability. Firstly, most studies have focused on the correlations with pedestrian environmental factors. Some of these studies have been used as decision supporting tool for pedestrian-related policy (Koschinsky et al. 2017; Gilderbloom et al. 2015; Azmi & Karim 2012; Frank et al. 2005; Leslie et al. 2005). For example, Greenwald & Boarnet (2001) explained the urbanism-neotraditional paradigm that investigates walking trips. In addition, the introduction of transit-oriented development (TOD) raises attention to walking benefits and promotes the improvement of individual physical health and environmental quality.

The second approach in previous walkability studies is that of the activity-based approach. It states that behaviour not only strongly depends on environmental factors, but that there are more forces which influence pedestrian behavioural choices in order to reach destinations and participate in every-day activities (Tribby et al. 2015; Jun & Hur 2015; Lindelöw et al. 2014; Dijst et al. 2002). For example, the mobility to work of someone without a car might predominantly be affected by the commuting distance. However, perhaps sidewalk facilities and external variables, such as weather conditions, may be critical as well for someone performing non-mandatory activities (such as walking for exercising or sightseeing).

Most walkability studies focused on environmental factors such as the land-use diversity index, the street intersection density, and the presence of public transit stations. They find that the attribute 'local opportunities' is selected as a significant variable in affecting the pedestrian walkability. Kim et al. (2014) estimated the effects

of the built environment on the pedestrian satisfaction. They built a multi-level model using an ordered logit regression model differentiating meso-scale (zonal) and micro-scale (personal) environmental factors. This approach is appropriate to understand the underlying influence of pedestrian environment quality on the walking score, such as the walkable distance from points of interest. However, all previous studies lack a reliable data collection and proper analytics might offend a lot of people. With respect to this context, this study intends to also apply disaggregated land-use features.

The main purpose in this study is to establish which kind of attributes influence the pedestrians' satisfaction. A pedestrian walkability model was developed which identifies the determinant factors influencing the level of pedestrian satisfaction. Various data from different sources were collected, including environmental and personal flavoured information. They were categorised into five distinctly themed variable groups. Discrete choice models were trained on the data to investigate correlations with pedestrian satisfaction. Additionally, this study provides a more in depth analysis of the following elements: the personal variability, the individual's spatial cognition level and specific travel-purpose, and urban environmental factors.

The paper is structured as follows. The next section summarizes the theoretical background, which provides details of conventional pedestrian walkability research. Afterwards, the research setting and methods are presented, followed by the analysis results of the pedestrian satisfaction MNL models. More specifically, this section reports the analysis results and discusses key findings related to personal variability. Having estimated the most important factors, a more in-depth analysis was performed on the respondent's level of spatial cognition (strangers vs. regular visitors) and travel purpose (utilitarian vs. personal). In addition, the effects of the built-environment on the walkability were detailed by means of an analysis based on the composition of building

types and on the degree of pedestrian crowdedness. This analysis revealed clear evidence of the importance of personal variables in future walkability studies. Various land-use factors were included in the models. The final section summarizes the results and limitations, and suggests topics for future research.

Theoretical background

Walking behaviour research (such as predicting the movement of pedestrians) is not only valuable for transportation planning, but also for urban planning, land-use and traffic operational research (Antonini et al. 2006). H.J.P. Timmermans (2009) introduced pedestrian behavioural researches and their challenges towards data collection for measuring the pedestrian dynamics. Several issues emerged when trying to understand pedestrian walking behaviour such as for instance: measuring the pedestrian flow and volume, modelling pedestrian movement, and estimating the sidewalk LOS.

The goal of walkability research was initially to understand pedestrian behaviour. Recently the scope has extended; most walkability studies follow the conceptual idea of the '3Ds': density, diversity, and design (Cervero & Kockelman, 1997). Cervero et al. (2009) extended this idea with two additional variables: 'distance to transit' and 'destination accessibility'. The result, the '5Ds' concept, implements all built-environment variables. Previous walkability studies naturally accepted this approach, and exclusively used land-use factors when examining the significant factors that affect pedestrian satisfaction. More specifically, they focused on the effects of built-environment features, using GIS and space syntax methods (Azmi & Ahmad, 2015; Kim et al. 2014; Lee & Seo, 2013; Gebel et al. 2011; Manaugh & El-Geneidy 2011; Sundquist et al. 2011; Baran et al. 2008; Frank et al. 2005; Frank et al. 2007). Those studies commonly collected zonal socioeconomic characteristics and spatial data first, and subsequently merge these into one complete dataset. Afterwards, they examined the significant variables using discrete choice modelling techniques such as ordered logistic or multinomial logistic (MNL) regression models, or conventional linear regression models.

Different academic fields continue to conduct walkability studies based on the various definitions of walkability, approaching the concept from different point of views. In particular, American walkability studies (Gilderbloom et al. 2015; Manaugh & El-Geneidy, 2011; Baran et al. 2008; Frank et al. 2007; Frank et al. 2005) estimated not only the pedestrian satisfaction, but also involved criteria of regional urban design and the destination accessibility, as well as the distance to points of interest (POIs). By contrast, environmental health researches focused on the personal variability and considered the interrelationship with urban design or streets (Gebel et al. 2011; Sundquist et al. 2011; Van Dyck et al. 2011; Saarloos et al. 2009; Leslie et al. 2007). In these studies, environmental effects were investigated with the goal of improving personal physical activity of inhabitants.

Walking behaviour depends on personal preference and is the result of individual decisions (Hoogendoorn & Bovy, 2005). Walking experiences from travellers affect future behaviour: positive experiences may mean that similar decisions are likely to be chosen again next time, and that the new behaviour is sustained as a habitual behaviour (Kim et al. 2014; Ettema et al. 2011). Additionally, different environmental attributes have a different spatial extent of influence on behavioural choices of different individuals, similar to the different mobility ranges between cyclists and car owners. (Saarloos et al. 2009; Bhat & Guo, 2007).

Contrary to the main research streams, some studies focus on the effect of personal and contextual variables (external factors, such as weather etc.) on pedestrian walkability. They provided evidence of the importance of contextual characteristics on pedestrian satisfaction. Osaragi (2004) suggested that the pedestrian walking behaviour is in part affected by 'mental stresses' caused by a large amount of conditions to consider, such as the shortest path criteria, the perception of the environment, and

occasional events. Owen et al. (2007) stated that conventional walkability studies examining the relationship between the pedestrian environments and walking fail to control for the neighbourhood self-selection bias. It is argued that individuals self-select their neighbourhoods based on their underlying preference for activities. Additionally, Van Dyck et al. (2011) argued that person characteristics are relevant when evaluating walkability, but also land-use variables, such as housing prices. Lee (2013) examined the factors affecting walkability by means of a CHAID analysis. The results show that the time of the survey (morning or afternoon) and the previously used transportation mode (public transit, walking or automobile) were significant factors explaining the level of pedestrian satisfaction.

In summary, the degree of pedestrian satisfaction is employed as the indicator of pedestrian walkability in this study. Contrary to previous studies, this study regards the pedestrian walkability as the individual psychological emotion based on preceding travel-experiences during which information was acquired on, among others, walking facilities and environmental characteristics. As opposite to previous empirical pedestrian walkability models, this study tries to estimate the determinant factors that influence the level of pedestrian satisfaction. Similar to previous studies, this study also considers walking environment and pedestrian facilities.

The corresponding baseline of these definitions for walkability in this study is as follows: the walking satisfaction is the base of the walkability index (or score) and it is introduced to evaluate to what degree streets or places are comfortable for walking. In addition, the conceptual idea of walkability may be extended by the destination walking accessibility, i.e., distance to the nearest opportunity and the number of opportunities within a particular area (Delafontaine et al. 2012). This elaborates on the conceptual idea that a good walkability solely implicates a high level of comfort.

Data and Methods

Seoul Pedestrian Survey

The 'Seoul Pedestrian Survey' was conducted from September to November in 2009 and covered the whole Seoul area, which is the capital city of the Republic of Korea. It is a highly dense city with almost 10 million inhabitants (Statistics Korea, 2015). There are several methods to study pedestrian walkability, for example questionnaire surveys (City of Vancouver, 2013) or alternative methods such as multilayer laser-scanning and CCTV in pedestrian safety studies (Kim & Kim, 2011; Greene-Roesel et al. 2008).

There were 2 stages in the survey data collection; on one hand an interview was taken from respondents, on the other hand the pedestrian flow volumes per hour were counted. At each survey location, surveyors collected responses from 24 pedestrians on Tuesdays, Wednesdays, and Fridays, yielding 72 responses per location. Up to 2,400 individuals per day were interviewed (Seoul Metro Government, 2010). The questionnaire collected the level of pedestrian satisfaction when walking through the survey area, as well as the pedestrian's activity-travel and personal information. There were over 1,170 survey locations (see Figure 1), totalling 83,291 responses.

[INSERT FIGURE 1 HERE]

Datasets

The dataset used in this study contained attributes on five distinct themes (see Figure 2; a list of all variables and their definition can be found in Appendix):

(1) *Personal* attributes such as sociodemographic and other surveyed characteristics, in particular pedestrians' trip-purpose and frequency with travel party information, collected for each individual at all survey locations. When discussing personal variables, this study focused on the visiting frequency attribute, because it represents the personal variation in spatial cognition level. Frequent visitors and newcomers have an 'information gap' caused by the difference in acquired built-environment information, which may impact the degree of personal satisfaction. In addition, trip purpose, such as utilitarian (commuting, school, and other work-related activities) and recreational (culture, leisure, and other social activities) have a different correlation based on the previous studies (Kim et al. 2014; Gebel et al. 2011; Bhat & Guo, 2007). This study investigates the effects of those variables on the pedestrian satisfaction.

(2) *Facility* attributes which describe the walking facilities in the immediate environment of the survey location (within a 50m range). It involves sidewalk facilities, such as the presence of central lines, obstacles or fences. It also includes the presence of public transit: subway and bus stations as amenities of urban street design.

(3) *Pedestrian volume* attributes represent pedestrian volumes and crowdedness estimates which usually occur at the survey location. The pedestrian volume at each survey point was measured by the survey investigators using hand-held counters (i.e. clickers) during 5 min every hour, between 7AM and 9PM at weekdays (Seoul Metro Government, 2010). This data is therefore known for all the survey spots (variables V_07_08 to V_20_21). One intuitively can imagine that absolute pedestrian volumes may influence pedestrian satisfaction, both for regular as for infrequent visitors. However, one may also suspect some sort of

learning effect: perhaps regular visitors get accustomed to a certain level of pedestrian volume at a certain location and time. To investigate this effect, crowdedness factors were calculated. They were estimated by dividing the pedestrian volume at the time of the survey period by the average of pedestrian volume over a larger time window. We defined it on two different temporal levels:

i.   Considering pedestrian volumes for four time periods in a day (Equation 1): this coefficient can be understood as "right now it feels more (or less) crowded than during the rest of the day"

ii.   Considering whole-day pedestrian volumes for four days of the week (during which the survey was conducted) (Equation 2): this coefficient can be understood as "today feels like a really crowded (busy) day compared to the other days of the week (or the other way round)"

$$C_{stp} = \frac{PV_{stp}}{Avg(PV_{tp1}, PV_{tp2}, PV_{tp3}, PV_{tp4})} \tag{1}$$

$$C_{swd} = \frac{PV_{swd}}{Avg(PV_{wd1}, PV_{wd2}, PV_{wd3}, PV_{wd4})} \tag{2}$$

where, PV is the pedestrian volume; C_{stp} is the crowdedness at a specific survey time period; C_{swd} the crowdedness at a specific survey week day; tp and wd are temporal level parameters indicating 'time period' (within the survey day) and 'weekday' respectively. One day is divided in four distinct time periods as the exact time of surveying a participant is not known (available categories

are: 8-11AM, 11AM – 2PM, 2-5PM and 5-8PM). The surveys were conducted on four days of the week. All figures are location dependent.

Both crowdedness definitions take a value of 1 when the pedestrian volume is not out of the ordinary, i.e. when it is equal to the average volume. A value lower or higher than one indicates, respectively, a lower or higher than usual volume. The currently defined crowdedness factor only depends on volumes, not the width of the sidewalk, and may therefore not be interpreted in 'absolute' terms of being crowded or not. It should rather be interpreted in a relative fashion, i.e. more or less crowded / busy than compared to the average pedestrian volume.

Both indicators are included as explanatory variable in the stepwise multinomial logistic regression models. The weekday-based definition of crowdedness has no significant effect on walkability satisfaction and was not included in the final model by the stepwise selection feature. It is suspected that the fact that only weekdays were considered when conducting the survey (and no weekend days) causes this effect.

(4) *Land-use* attributes which reveal the land-use mix within a radius of 300 meters of the survey location. In order to compose the land-use datasets, first an influence zone consisting out of all census tracts within a range of 300 m of the survey location was defined for each of the survey locations (see Figure 3). Other Korean empirical pedestrian studies (Kang, 2015; Kim et al. 2014; Kim et al. 2013) generally assumed similar ranges (300-500m) for each survey point's buffer zone. Secondly, data from a number of government open-data platforms, such as geospatial datasets from the National Statistical Geographic Information Service (SGIS), Statistics Korea in 2012 and 'Seoul Pedestrian Survey' data

from the Seoul Open Data Plaza, Seoul Metro Government in 2012 were combined into one large dataset. The completed data was geospatially pre-processed. For instance, building types were classified into residential, commercial, public services, and so on. The gathered land-use data contains almost 0.8 million buildings, and was categorized in 8 types and specific floor area classes. The land-use mix was calculated based on the geospatial datasets in a land-use entropy index (Cervero & Kockelman, 1997). This index quantifies the level of mixed land-use based on the building-types. The value has a range from 0 to 1. The closer to 0, the simpler and monotonous the land-use, whereas a value closer to 1 represents a mixed land-use (Kang, 2015). The land-use entropy value is calculated based on the following formula (Equation 3):

$$Mix\_i_{sp} = \frac{\sum_{i=1}^{n} P_i ln(P_i)}{ln(n)} \tag{3}$$

where, n is the number of land-use types (using seven building types), P_i is the proportion of each land-use i, sp represents each survey point.

(5) *Weather*-related attributes for 3 months collected by Korea Aviation Meteorological Agency were merged using the survey date. In this research, the weather information was inserted as an external contextual situation. It is plausible that it indirectly impacts the individual's subjective state when walking through the survey area. For example, pedestrians might be less satisfied when it rains or in case of a very high or low temperature or humidity. By contrast, a nice warm cloudless day might positively affect an individual's walking experience.

[INSERT FIGURE 2 HERE]

[INSERT FIGURE 3 HERE]

The independent variables of the different themes have different spatial scales. In fact, two different geographical levels were included, consisting of the survey location layer (personal, pedestrian volume) and the census tract layer (facilities, land-use, and weather data).

The dependent variable in this research is the pedestrian satisfaction (walkability). Individuals that completed the survey were asked to rate their overall satisfaction (related to walkability) with a score ranging from one to five; i.e. a five-level Likert scale (see Figure 4).

[INSERT FIGURE 4 HERE]

An indicated walkability score of three represented a neutral satisfaction and a score of one and five respectively indicated a very unsatisfactory or very satisfactory walking experience. Within the current research, these responses were aggregated into three instead of five possible outcomes: a negative, neutral, or positive walking satisfaction. This transformed walkability attribute with three discrete outcomes is used as the dependent variable in the pedestrian walkability model (see Figure 2). The advantage of this transformation is that the model is easier to discuss and that some statements can be made with more confidence. As can be seen in Figure 4, the number of observations for

the "very unsatisfactory" alternative is quite low. The aggregation also has the advantage of increasing the number of observations for each alternative, mediating this issue of few observations. It is in the scope of this research to identify the factors that influence walkability, either positively or negatively. If a model was built using five alternatives for pedestrian satisfaction, it is more likely to contain less significant coefficient estimates than is the case for a model with only three discrete outcomes. In other words; within the current research it is more enlightening to be able to claim with confidence that a property (from one of the variable themes) negatively influences pedestrian satisfaction, rather than being able to make the distinction between "very unsatisfactory" or "unsatisfactory" based on less significant estimates. This justifies the aggregation from five into three alternatives, thereby removing some of the richness of the data.

Modelling: MNL model

The discrete choice model used in this research is the Multinomial Logit (MNL) model. The statistical software package SAS 9.4 offers this functionality in an easy to use procedure (proc logistic). An important property of the logistic regression model is that always one of the dependent variable outcomes serves as the reference category. This is of importance when discussing the estimated coefficients of the model. There are three discrete nominal alternatives in the dependent attribute (the walkability satisfaction): a negative, neutral or positive response. The alternative of a neutral walking experience was chosen as the reference, and therefore only coefficient estimates were calculated for a negative- and for a positive walking experience, both with respect to the common reference alternative: a neutral walking experience. This model structure is according to standard discrete choice modelling and is fully automated by SAS (Equation 4). The dependent attribute (pedestrian walkability) is used in a logit regression that takes the

form:

$$Logit(z) = \sum \beta_{P(i)} \cdot A_{P(i)} + \sum \beta_{F(j)} \cdot A_{F(j)} + \sum \beta_{PV(k)} \cdot A_{PV(k)}$$

$$+ \sum \beta_{LU(l)} \cdot A_{LU(l)} + \sum \beta_{W(m)} \cdot A_{W(m)} + \varepsilon_{(z)}$$

(5)

where, $A_{<>}$ represent the attribute and $\beta_{<>}$ the corresponding estimated coefficients. *P* stands for the personal-themed attributes, *F* for the facility-themed ones, *PV* for the pedestrian volumes variables, *LU* for the land-use attributes and *W* for the weather-related information. Some of the attributes in these datasets are categorical (nominal) in nature rather than continuous (numeric). In order to correctly take into account their effects, dummy variables were created and used in the models. This process is an automated feature in SAS.

According to the recent literature, the latent class framework may also be used to investigate heterogeneity in travel behaviour (McFadden and Train 2000). A latent class model allows to identify the best number of segments and the membership model of different segments (Walker and Li 2007). Within the latent class framework, variable effects are estimated separately for each segment under the assumptions that individuals' preferences within the same segment are homogeneous (Feng, Arentze, and Timmermans 2013).

Kim et al. (2014) were the first to applied the latent class framework in a pedestrian walkability case study. However, this study was limited by only investigating trip purpose consisting of utilitarian or personal nature. The current study focused on the examination of *what* indicators influence the walkability. Future research might

apply the latent class approach, which focuses on the heterogeneous effects of pedestrian (individual) contexts on their satisfaction level.

The complete dataset of more than 80,000 surveyed individuals was split into a training (75%) and a test (25%) set according to a random selection in order to allow for cross-validation. To assess the relative predictive power of the five attribute categories, a distinct multinomial logistic regression model was estimated based on the five distinct sets of dependent variables. These models were fitted using the training datasets. The test set was used to check the model validity on new data in a cross-validation.

For measuring the parameters, a stepwise selection feature was used. A significance level (p-value) of 10% is required to allow a variable into the model, and a significance level of 15% is required for a variable to stay in the model. The threshold levels were chosen a bit more generous than more stringent common levels of 5% or 10%. This stepwise selection feature was employed to exclude rather insignificant attributes, reducing the model size. All five models used the same threshold levels.

In order to compare the model strength, the adjusted generalized coefficient of determination (referred to as adj. gen. R^2) figure was requested in SAS. This is sort of a pseudo R-square based on log-likelihood with similar properties as the R-square coefficient used in non-discrete choice models.

Examining the final model (see also the Model Estimation section and Table 1), one might notice some coefficient estimates that exceed the aforementioned threshold levels. It is a widely accepted practice to include a significant attribute even if one (or more) estimate of an alternative from this attribute is not significant.

[INSERT TABLE 1 HERE]

Analysis results

Model estimation

Table 2 lists the fit characteristics for both train and test set. As can be observed, rather low adj. gen. R-square values are obtained (similar as in, for example, Kim et al. 2014) indicating that there is significant variability in the walkability score that is not explained by the models. However, as will be shown afterwards, the models revealed extremely significant relations between the independent attributes and the walkability score, despite the large fraction of unaccounted variability in reported walkability score. Since uncovering these relations is the prominent point of this research and not making predictions, the absolute magnitude of the adj. gen. R-square value can be safely ignored. Having included the most common attributes of all five-attribute categories, the relative predictive strength can be evaluated with confidence. The model validity is confirmed by the good correspondence between train- and test set fit characteristics.

A comparison of the adjusted R-square values in Table 2 shows that the group of *personal* attributes yields the most explanatory power (0.06), followed by the closely matched groups of *land-use* (0.030) and *facility* (0.027) attributes. Rather surprisingly, the *pedestrian volumes* (0.012) and *weather*-related attributes (0.008) account for a much lower fraction of the reported walkability score than the earlier mentioned groups.

These models were fitted using the training dataset. The test set was used to check the model validity on new data as in a cross-validation approach, and these are also the figures mentioned above. One observes that the reported adj. gen. R^2 values for training- and test set are very close to each other, indicating that there is no serious case of overfitting. Table 2 also lists the likelihood ratio test results. All models are significantly better (sig < 0.001) than the null hypothesis model (intercepts only).

Comparing these test results relative to each other yields the same conclusion as previously formulated based on the generalized R-square value. The error rates listed in Table 2 are those in case the models were to be used as predictive models, which is however not the intend of this paper. The error rates are relatively high, since not all variability of the pedestrian satisfaction is controlled by the model as was also revealed by the adj. gen. R^2 values.

The fit characteristics for a model containing *all* themes of variables are also listed in Table 2 in addition to the more detailed results in Table 1. This model was constructed using the stepwise procedure. Afterwards, some numeric attributes having large correlations with others were removed to limit unwanted correlations between the explanatory variables, yielding more stable coefficient estimates. For example, an undesirable level of correlation was present between some of the volume-themed variables, as well as between some of the weather-related attributes. After remediating most of the correlation issues by excluding a number of these variables, the model was refitted, yielding more reliable coefficient estimates.

As revealed above, the model estimation results indicate that personal and contextual attributes most significantly influence the pedestrian walkability. In the next sub-section, the heterogeneity of pedestrian walkability according to respondents' perceived spatial cognition level (i.e. regional familiarity) and trip purpose is investigated. This subsequent work shows that the understanding of personal diversity is crucial to understand pedestrian walkability.

[INSERT TABLE 2 HERE]

Spatial cognition level

In this section the effect of the spatial cognition level is investigated. It is hypothesized that frequent visitors (residents) might have a different rating of the walking experience compared to first-time visitors (e.g. tourists). For example: perhaps first-time visitors rate the effect of their direct walking environment (sidewalks, obstacles) more highly than regular visitors who are already well familiar with this setting. In an attempt to reveal the effect of being familiar with the environment, two (independent) tests were performed as follows:

In the first test, two subsets of participants were created. The first contained individuals who reported never before having visited the survey location; the second contained participants who live in the same administrative district as the survey location and who visit this location at least 1-2 times per week. This first test is similar in construct as the one summarized in Table 2, but the models were fitted on two *subsets* of the data (frequent and infrequent visitors respectively). Five new models were estimated (one for each of the variable themes) on both of these two subsets. Subsequently, the relative magnitude of the generalized R-square values of these models was analysed. From this comparison, it could be concluded that first-time visitors and regular visitors employ similar criteria to rate their walking experience: for both of these strata the highest explanatory power comes from *personal* attributes, followed by the closely matched *facility* characteristics and *land-use*-themed variables. This is similar to the findings of the previous section.

The second test investigates the logistic regression coefficients from the complete model. Figure 5 shows the estimated coefficients versus the visiting frequency. Several conclusions can be formulated. The probability on a negative as well

as a positive walking satisfaction increases with increasing visiting frequency. This indicates that frequent visitors are more critical and more often report a non-neutral rating. There seems to be a bias towards a positive walking experience for frequent visitors, which may be intuitively be explained. A remarkable exception to this trend can be observed for non-general visitors as the category "1-3 days per 6 months". This observation seems not to be caused by insufficient sample size or correlation to other attributes. The effect may be investigated in detail in future research.

This result is similar to previous travel behaviour studies for estimating the personal variability of activity-travel patterns and decision processes. Joh et al. (2011) examined the impact of acquired (provision) information using a CHAID analysis. They found that contextual variables, work status, monthly income and time of day significantly influence the individual's decision. In addition, Moiseeva et al. (2014) analysed the variability of activity-travel patterns of newcomers during 3 months. The results indicate that interpersonal variability is significantly higher than intrapersonal variability. Activity-travel patterns may correlate with sociodemographic characteristics such as gender and country of origin.

The spatial cognition level greatly differs between regular visitors (daily visit) and strangers (first-time visit). Regular visitors have the largest coefficients for both a negative and positive walking satisfaction compared to any other group (except infrequent visitors in the category of "1-3 days for 6 months"). Strangers are characterised with the lowest coefficient estimate values in the model.

[INSERT FIGURE 5 HERE]

Travel purpose

Travel purpose (or activities) is widely used in transportation demand models. In activity-based models, it is hypothesized that activities are at the foundation of the need for transportation. Having such major importance, the effect of travel purpose on pedestrian satisfaction is investigated in this section.

Figure 6 summarizes the logistic regression coefficients for the travel purpose. The effect of travel purpose on the pedestrian satisfaction is highly significant. Compared to the reference class (the travel purpose "home"), most other travel purposes have a significantly different effect on pedestrian satisfaction. Almost all travel purpose coefficients are positive, indicating that pedestrians are more critical or stressed and have a higher probability to report a non-neutral walkability rating than in the case of the reference "going home".

Utilitarian walking with travel purposes such as "commuting", "business", "bank/post/office", and "transfer between public transport modes" tend to imply a more negatively impact (coefficient) than personal walking activity-types such as "personal" walking, "shopping", "social & leisure". These activities have a positive impact on pedestrian satisfaction. This experimental result may be related to the conceptual idea of 'mental stress', which have been shown to be linked to pedestrian attributes and environmental conditions in Osaragi (2004).

A second logistic regression model was fitted in which the travel purpose attribute was modified. The travel purposes "commuting", "business", "school", "education", "bank/post/office", "bring & get" and "public mode transfer" were aggregated into a class of "*utilitarian*" travel purposes. The remaining travel purposes, i.e. "personal", "shopping", "social/leisure" and "home" activities were grouped into the

"*personal*" class of travel purposes.

From this second model, it was possible to conclude that travel purposes of *personal* nature are less likely to accompany a non-neutral walkability satisfaction than travel purposes of *utilitarian* nature. Possibly some form of stress resulting from the external obligation of utilitarian travel purposes is at the base of this behaviour, but this remains to be confirmed in other research.

[INSERT FIGURE 6 HERE]

Environmental effects

Based on the conventional walkability studies, environmental characteristics are the major determinant factors which influence the pedestrian walkability. In this study also the effects of environmental factors such as crowdedness, building-types, and residential location were investigated. First, the effect of crowdedness, the higher- or lower-than-usual number of pedestrians, has been investigated by including two figures in the logistic regression model (see also *Pedestrian volume* in the *Data and Methods* section). The crowdedness based on within-day volume changes was highly significant when explaining pedestrian satisfaction. An increasing crowdedness results in an increased probability on a negative walkability experience and a decreased probability on a positive one (and vice-versa). This is consistent with the hypothesis that higher than usual pedestrian volumes may offer some form of discomfort. Table 3 lists the regression coefficients. In case that for example the pedestrian volume is double that of the day's average, then the coefficient estimates imply that the odds on a negative

walkability experience (compared to a neutral one) increase by almost 20%. At the same time, the odds for a positive walkability experience will decrease by 10%.

[INSERT TABLE 3 HERE]

Secondly, some significant influences of land-use on pedestrian satisfaction were found in this research. Figure 7(a) shows some of these influences. One noteworthy finding is that both the probability on a negative as a positive walking experience increase with increasing educational buildings in the predefined buffer zone. Perhaps this is the consequence of a different sociodemographic composition, or perhaps people are more critical when children's safety or well-being is affected. Industrial buildings result in a higher probability on a low pedestrian satisfaction and a lower probability on a high level of pedestrian satisfaction. The presence of cultural buildings tends to increase the probability on a positive walkability experience. Commercial buildings tend to increase the probability on a neutral pedestrian satisfaction.

Figure 7(b) indicates the influence of building floor area for several land-use types. Remarkable is that the estimated coefficients related to industrial floor area are much larger (in absolute value) than those related to other buildings types. An increasing floor area of industrial buildings within the buffer zone yields a higher probability on a positive walking experience and a lower probability on a negative walking experience, which seems counterintuitive.

In addition to the discussion above, the effect of the land-use mix was analysed. The land use mix was analysed in the form of an entropy value, as detailed in the Data

and Methods section. The logistic regression dictates that a higher entropy results in lower probabilities for both the positive as the negative walking experiences, and therefore implies a higher probability on a neutral experience. A low entropy value represents a monotonous land-use, which might be a reason for a negative, or in other cases positive (e.g. shopping streets) walking satisfaction (and therefore a lower probability on a neutral satisfaction). If a low entropy lowers the probability on a neutral satisfaction, a high entropy per definition increases the probability on a neutral satisfaction.

[INSERT FIGURE 7 HERE]

Note that some attempts were made to prevent poor coefficient stability because of high correlation between the numeric explanatory attributes (multicollinearity), but some minor correlation may not be avoided.

In the group of *personal*-themed variables, the variable describing the residence location is one of the most significant ones. There is some significant differentiation, either negatively or positively, between walkability satisfaction levels of different residential areas. The logistic regression coefficients are shown in Figure 8. A high coefficient for a negative walking experience was found in Gangdong-*gu*, a positive walkability occurred in Gangnam-*gu* and Seodaemun-*gu*. Perhaps this finding might be related to the urban scene of a location. This attribute of residential area possibly captures some variation in pedestrian satisfaction caused by personal or land-use characteristics which were not surveyed.

A variable indicating whether the respondent was surveyed in the same

administrative district as that of his residence (P_REGION) was also highly significant. Individuals being interviewed in the district where they live have a lower probability on a low pedestrian satisfaction and a higher probability on a high level of satisfaction. This may be explained in part by their spatial cognition level, but perhaps also some form of emotional affinity with the living area is behind this effect.

[INSERT FIGURE 8 HERE]

Conclusion

This study investigated the significant factors influencing the pedestrian satisfaction. Moreover, a pedestrian walkability model was built using discrete choice modelling. The datasets were compiled from the 'Seoul Pedestrian Survey', which was collected by surveyors, and from geospatial datasets proving by the governments' open data platform. The main purpose of this study was to find the relative impact of five differently-themed sets of attributes: personal, walk-facilities, land-use, pedestrian volumes, and weather-related variables. The key findings were explored, such as the effects of personal variability and environmental factors.

Five partial pedestrian walkability models were built using MNL regression models. Regression coefficients were estimated and the adjusted generalised R-square values were used to compare five distinct variable-themes. The results showed that *personal* attributes represent the highest explanatory power; higher than the environmentally themed *walk-facilities* and *land-use* attributes. These results provide additional evidence for previous pedestrian behaviour research challenges. Following the activity-based transportation approach, personal variability is the most determinant factor to impact the pedestrian satisfaction. However, *walk-facilities* and *land-use* variables also have a significant influence on the pedestrian walkability.

One model was built starting from all variables in the dataset. Variability and environmental effects such as the effects of crowdedness, building types, and respondents' residential places were discussed based on this model. To estimate the effect of personal information, spatial cognition was investigated. It was estimated based on the degree of visitors' previous walking experiences on the survey location, categorized into frequent and first-time visitors. Different, significant logistic regression

coefficients were found. Frequent visitors were associated with high coefficient values, both for the negative and positive walkability alternatives, compared to first-time visitors. In addition, when comparing the effects of specific activity-types, classified into utilitarian and personal travel purposes, personal walking purposes have been found responsible for more positive walkability than utilitarian walking. This is possibly related to the individual's 'mental stresses' from the external obligation of utilitarian walking.

Several environmental effects were investigated. Firstly, the effects of crowdedness based on pedestrian flow volumes at each survey location during specific time-periods were analysed. The crowdedness factors are calculated on two temporal levels: based on survey-time periods and on whole-day volumes respectively. The results show that an increased crowdedness increases the probability on a negatively pedestrian satisfaction. Secondly, the effects of building type were verified. Cultural building types tend to increase the probability on a positive walkability. Commercial buildings tend to increase the probability on a neutral pedestrian satisfaction. Previous walkability studies, which focused on the roles of built-environmental factors, also found these variables to significantly impact the pedestrian satisfaction. Future research should thoroughly check the effect of land-use mix (entropy value) and estimate the hidden impacts, such as neighbourhood effects, using multilevel modelling methods. In the last part of this research, the effect of residence location was investigated. It was based on the *personal*-themed variables, describing the residence location of each Seoul administrative districts (*Gu*). In further research, this neighbourhood effects could be investigated in more detail using multilevel models, such as multilevel MNL or latent class regression models.

This study confirmed the importance of personal variability in pedestrian

satisfaction using a large amount of data for development pedestrian walkability models, employing a cross-validation process for model validation. The contributions of this study are as follows: first, the current study provides evidence that personal and contextual variables are important factors of the pedestrian walkability. Secondly, the personal variability, travel experience and walking purpose directly impact the level of pedestrian satisfaction. Finally, the results of this study provide comprehensive information in a set of individual characteristics, which affect pedestrian satisfaction.

References

Antonini, G., Bierlaire, M., & Weber, M. 2006. Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, *40*(8), 667–687. doi:10.1016/j.trb.2005.09.006

Azmi, D. I., & Ahmad, P. 2015. A GIS Approach: Determinant of Neighbourhood Environment Indices in Influencing Walkability between Two Precincts in Putrajaya. *Procedia - Social and Behavioral Sciences*, *170*, 557–566. doi:10.1016/j.sbspro.2015.01.057

Azmi, D.I. & Karim, H.A., 2012. Implications of Walkability Towards Promoting Sustainable Urban Neighbourhood. *Procedia - Social and Behavioral Sciences*, 50(July), pp.204–213. doi: 10.1016/j.sbspro.2012.08.028.

Baran, P.K., Rodríguez, D. a. & Khattak, A.J. 2008. Space Syntax and Walking in a New Urbanist and Suburban Neighbourhoods. *Journal of Urban Design*, 13(1), pp.5–28. doi: 10.1080/13574800701803498.

Bhat, C. R., & Guo, J. Y. 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, *41*(5), 506–526. doi:10.1016/j.trb.2005.12.005

Cervero, R., & Kockelman, K. 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219. doi:10.1016/S1361-9209(97)00009-6

Cervero, R., Sarmiento, O. L., Jacoby, E., Gomez, L. F., & Neiman, A. 2009. Influences of Built Environments on Walking and Cycling: Lessons from Bogotá. *International Journal of Sustainable Transportation*, *3*(4), 203–226. doi:10.1080/15568310802178314

City of Vancouver. 2013. *Pedestrian Volume and Opinion Survey*.

Delafontaine, M., Neutens, T. & Van de Weghe, N., 2012. A GIS toolkit for measuring and mapping space–time accessibility from a place-based perspective. *International Journal of Geographical Information Science*, 26(6), pp.1131–1154. doi: 10.1080/13658816.2011.635593

Dijst, M., Jong, T. de, & Eck, J. R. van. 2002. Opportunities for transport mode change: an exploration of a disaggregated approach. *Environment and Planning B: Planning and Design*, *29*(3), 413–430. doi:10.1068/b12811

Ettema, D., Gärling, T., Eriksson, L., Friman, M., Olsson, L. E., & Fujii, S. 2011. Satisfaction with travel and subjective well-being: Development and test of a measurement tool. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(3), 167–175. doi:10.1016/j.trf.2010.11.002

Frank, L. D., Saelens, B. E., Powell, K. E., & Chapman, J. E. 2007. Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Social Science and Medicine*, *65*(9), 1898–1914. doi:10.1016/j.socscimed.2007.05.053

Frank, L. D., Schmid, T. L., Sallis, J. F., Chapman, J., & Saelens, B. E. 2005. Linking objectively measured physical activity with objectively measured urban form. *American Journal of Preventive Medicine*, *28*(2), 117–125. doi:10.1016/j.amepre.2004.11.001

Gebel, K., Bauman, A. E., Sugiyama, T., & Owen, N. 2011. Mismatch between perceived and objectively assessed neighborhood walkability attributes: Prospective relationships with walking and weight gain. *Health and Place*, *17*(2), 519–524. doi:10.1016/j.healthplace.2010.12.008

Gehrke, Steven R., and Kelly J. Clifton. 2017. "A Pathway Linking Smart Growth Neighborhoods to Home-Based Pedestrian Travel." *Travel Behaviour and Society* 7: 52–62. doi:10.1016/j.tbs.2017.02.003.

Gilderbloom, J. I., Riggs, W. W., & Meares, W. L. 2015. Does walkability matter? An examination of walkability's impact on housing values, foreclosures and crime. *Cities*, *42*, 13–24. doi:10.1016/j.cities.2014.08.001

Greene-Roesel, R., Diogenes, M. C., Ragland, D. R., & Lindau, L. a. 2008. Effectiveness of a Commercially Available Automated Pedestrian Counting Device in Urban Environments : Comparison with Manual Counts. In *TRB Annual Meeting*, Washington D.C., USA.

Greenwald, M., & Boarnet, M. 2001. Built Environment as Determinant of Walking Behavior: Analyzing Nonwork Pedestrian Travel in Portland, Oregon. *Transportation Research Record*, *1780*(1), 33–41. doi:10.3141/1780-05

H. J. P. Timmermans. 2009. *Pedestrian Behavior: Models, Data Collection and Applications*. Emerald Group Publishing.

Hoogendoorn, S. P., & Bovy, P. H. L. 2005. Pedestrian travel behavior modeling. *Networks and Spatial Economics*, *5*(2), 193–216. doi:10.1007/s11067-005-2629-y

Hussein, Mohamed, and Tarek Sayed. 2017. "A Bi-Directional Agent-Based Pedestrian Microscopic Model." *Transportmetrica A: Transport Science* 13 (4): 326–355. doi:10.1080/23249935.2016.1266531.

Joh, C. H., Lee, B., Bin, M., Arentze, T., & Timmermans, H. 2011. Exploring the use of travel information - identifying contextual market segmentation in Seoul, Korea. *Journal of Transport Geography*, *19*(6), 1245–1251. doi:10.1016/j.jtrangeo.2011.06.001

Kang, C.-D. 2015. The effects of spatial accessibility and centrality to land use on walking in Seoul, Korea. *Cities*, *46*, 94–103. doi:10.1016/j.cities.2015.05.006

Kim, D., Ko, J. & Lee, Y., 2013. Estimating Pedestrian Traffic Volume: A Preliminary Analysis. In *The Eastern Asia Society for Transportation Studies*. Available at: http://easts.info/on-line/proceedings/vol9/PDF/P154.pdf.

Kim, S., Park, S. & Lee, J.S. 2014. Meso- or micro-scale? Environmental factors influencing pedestrian satisfaction. *Transportation Research Part D: Transport and Environment*, 30, pp.10–20. doi: 10.1016/j.trd.2014.05.005.

Kim, T. H., Kim, J., Park. E. K., & Kang, J. E. 2011. *A Study on Determining Factors on Pedestrian Volume by Station Area Types*. The Seoul Institute. (In Korean).

Lee, S. & Seo, K.W. 2013. Combining Space Syntax With GIS-Based Built Environment Measures in Pedestrian Walking Activity. In *9th International Space Syntax Symposium*. Seoul, South Korea.

Lee, W. Do. 2013. Identifying the Factors Affecting Pedestrian Flow Volume and Walkability Using the "Seoul Pedestrian Survey" Data. In *Computers in Urban Planning and Urban Management (CUPUM)*. Utrecht, the Netherlands.

Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. 2007. Walkability of local communities: Using geographic information systems to objectively assess relevant environmental attributes. *Health and Place*, *13*(1), 111–122. doi:10.1016/j.healthplace.2005.11.001

Leslie, E., Saelens, B., Frank, L., Owen, N., Bauman, A., Coffee, N., & Hugo, G. 2005. Residents' perceptions of walkability attributes in objectively different neighbourhoods: A pilot study. *Health and Place*, *11*(3), 227–236. doi:10.1016/j.healthplace.2004.05.005

Lindelöw, D., Svensson, Å., Sternudd, C., & Johansson, M. 2014. What Limits The Pedestrian? Exploring Perceptions of Walking in the Built Environment and in the Context of Every-Day Life. *Journal of Transport & Health*, *1*(4), pp 223–231. doi:10.1016/j.jth.2014.09.002

Manaugh, K. & El-Geneidy, A. 2011. Validating walkability indices: How do different households respond to the walkability of their neighborhood? *Transportation Research Part D: Transport and Environment, 16(4), pp.309–315. doi: 10.1016/j.trd.2011.01.009.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447–470. doi:10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

Moiseeva, a, Timmermans, H., Choi, J., & Joh, C.-H. 2014. Sequence Alignment Analysis of Variability in Activity Travel Patterns Through 8 Weeks of Diary Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2412*(January), 49–56. doi:10.3141/2412-06

Neutens, T., Versichele, M. & Schwanen, T. 2010. Arranging place and time: A GIS toolkit to assess person-based accessibility of urban opportunities. *Applied Geography*, 30(4), pp.561–575. doi: 10.1016/j.apgeog.2010.05.006.

Nurul Habib, Khandker, Xiao Han, and William Haoyang Lin. 2014. "Joint Modelling of Propensity and Distance for Walking-Trip Generation." *Transportmetrica A: Transport Science* 10 (5): 420–436. doi:10.1080/23249935.2013.778356.

Osaragi, T. 2004. Modeling of pedestrian behavior and its applications to spatial evaluation. In *the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Owen, N., Cerin, E., Leslie, E., duToit, L., Coffee, N., Frank, L. D., Bauman, A. E., Hugo, G., Saelens, B. E. and Sallis, J. F. 2007. Neighborhood Walkability and the Walking Behavior of Australian Adults. *American Journal of Preventive Medicine*, *33*(5), 387–395. doi:10.1016/j.amepre.2007.07.025

Papadimitriou, E., Yannis, G. & Golias, J. 2009. A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3), pp.242–255.

Saarloos, D., Joh, C. H., Zhang, J., & Fujiwara, A. 2010. A segmentation study of pedestrian weekend activity patterns in a central business district. *Journal of Retailing and Consumer Services*, *17*(2), 119–129. doi:10.1016/j.jretconser.2009.11.002

Saarloos, D., Kim, J.-E. & Timmermans, H. 2009. The Built Environment and Health: Introducing Individual Space-Time Behavior. *International Journal of Environmental Research and Public Health*, 6(6), pp.1724–1743.

Seoul Open Data Plaza. 2009. *Seoul Pedestrian Survey Data*. Retrieved January 13, 2012 from http://data.seoul.go.kr/

Seoul Pedestrian Survey Report. 2010. *Seoul Metro Government*.

Statistical Geographic Information Service. 2010. *Census tract-based Geospatial Data*. Retrieved July 9, 2012 from http://sgis.kostat.go.kr/

Sundquist, K., Eriksson, U., Kawakami, N., Skog, L., Ohlsson, H., & Arvidsson, D. 2011. Neighborhood walkability, physical activity, and walking behavior: The Swedish Neighborhood and Physical Activity (SNAP) study. *Social Science and Medicine*, *72*(8), 1266–1273. doi:10.1016/j.socscimed.2011.03.004

Van Dyck, D., Cardon, G., Deforche, B., Owen, N., & De Bourdeaudhuij, I. 2011. Relationships between neighborhood walkability and adults' physical activity:

How important is residential self-selection? *Health and Place*, 17(4), 1011–1014. doi:10.1016/j.healthplace.2011.05.005

Walker, J. L., & Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9(1), 77–101. doi:10.1007/s10109-006-0030-0

Xie, Siqi, and S.C. Wong. 2015. "A Bayesian Inference Approach to the Development of a Multidirectional Pedestrian Stream Model." *Transportmetrica A: Transport Science* 11 (1): 61–73. doi:10.1080/23249935.2014.924165.

Yin, Li, and Zhenxin Wang. 2016. "Measuring Visual Enclosure for Street Walkability: Using Machine Learning Algorithms and Google Street View Imagery." *Applied Geography* 76: 147–153. doi:10.1016/j.apgeog.2016.09.024.

List of tables

Table 1: Variable significance levels after the stepwise selection procedure for the model containing *all* themes of variables.

Theme	Variables	DF	Wald Chi-Square	Theme	Variables	DF	Wald Chi-Square
Personal	P_GENDER**	2	12.017	Land-use	L_T_POPULATION**	2	10.070
	P_SURVEY_DATE**	2	16.294		L_T_HOUSEHOLDS	2	4.866
	P_SURVEY_DAY	6	12.507		L_T_COMPANIES**	2	38.741
	P_AGE_CATEGORY**	10	83.073		L_T_EMPLOYMENT**	2	19.323
	P_RESIDENCE_DISTRICT**	50	1274.265		L_T_HOUSING**	2	11.785
	P_REGION**	2	33.746		L_B_RESIDENTIAL_N**	2	299.033
	P_VISIT_PURPOSE**	20	143.807		L_B_COMMERCIAL_N**	2	21.930
	P_VISIT_FREQ**	10	131.830		L_B_CULTURAL_N**	2	18.221
	P_VISIT_PARTY**	6	54.780		L_B_EDUCATIONAL_N**	2	34.700
	P_PREVIOUSLY_USED_MODE**	10	139.019		L_B_INDUSTRY_N**	2	52.223
	P_JOB**	18	110.755		L_B_ETC_N**	2	30.468
Facilities	F_WIDTH_SIDEWALK**	2	114.061		L_B_TOTAL_A*	2	7.359
	F_CENTRAL**	2	142.558		L_B_COMMERCIAL_A**	2	24.705
	F_NUM_LANES**	2	161.907		L_B_CULTURAL_A	2	1.888
	F_BUS_LANE**	2	34.686		L_B_OFFICE_A**	2	32.129
	F_OBSTACLE**	2	31.723		L_B_INDUSTRIAL_A**	2	56.386
	F_BRAILLE_BLOCK**	2	14.440		L_B_ETC_A**	2	20.917
	F_WALK_ROAD	4	188.324		L_B_ENTROPY**	2	12.443
	F_FENCE**	2	13.163	Weather	W_MAX_TEMPERATURE**	2	18.196
	F_SLOPE**	2	19.126		W_MAX_SEA_LEVEL_PRESSURE**	2	10.780
	F_BUS_STOPS**	2	29.081		W_MAX_VISIBILITY**	2	8.568
	F_CROSSING**	2	18.468		W_MEAN_WIND_SPEED	2	4.760
	F_SUBWAY*	2	7.690		W_CLOUDCOVER**	2	8.851
Pedestrian Volumes	V_07_08**	2	24.982		W_EVENTS**	8	22.758
	V_09_10	2	4.016				
	V_12_13**	2	40.439				
	V_14_15**	2	25.873				
	V_20_21**	2	12.972				
	V_CROWDEDNESS** (WITHIN-DAY_ DEFINITION)	2	66.452				

Note **: (p < .01), *: (p <. 05).

Table 2. Model estimation results (containing the adjusted generalized coefficient of determination).

		Personal	Facilities	Pedestrian Volumes	Land-use	Weather	All
Likelihood ratio (sig<0.0001)		3230	1672	881	1595	482	6069
train-set (75%)	adj. gen. R^2	0.058	0.030	0.016	0.029	0.009	0.105
	error rate	52.35%	55.20%	56.52%	54.63%	56.91%	50.83%
test-set (25%)	adj. gen. R^2	0.060	0.027	0.012	0.030	0.008	0.097
	error rate	52.14%	55.17%	55.62%	54.03%	56.65%	51.07%
# independent variables (before stepwise selection)		13	12	27	23	22	97
# independent variables (after stepwise selection)		10	12	18	16	14	53

Table 3. Logistic regression coefficients for the within-day crowdedness factor.

	C_{stp}* coefficient estimates	P-value
Negative walkability alternative	0.1703	<0.0001
Positive walkability alternative	-0.0997	0.0002

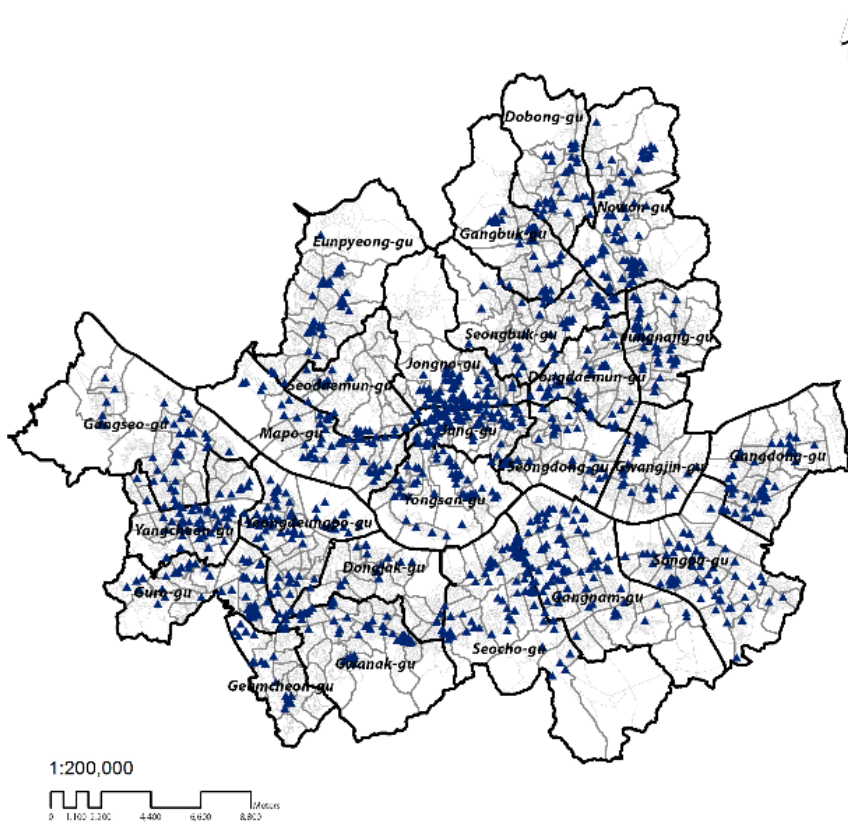Note *: Crowdedness for each survey time period.

List of figures

Figure 1. Seoul Pedestrian Interview Spots (total: 1,170).
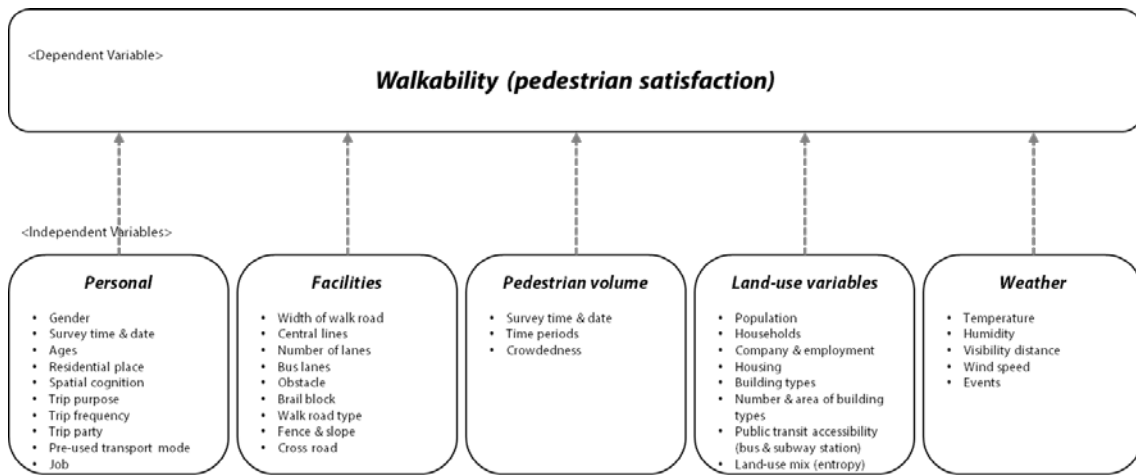
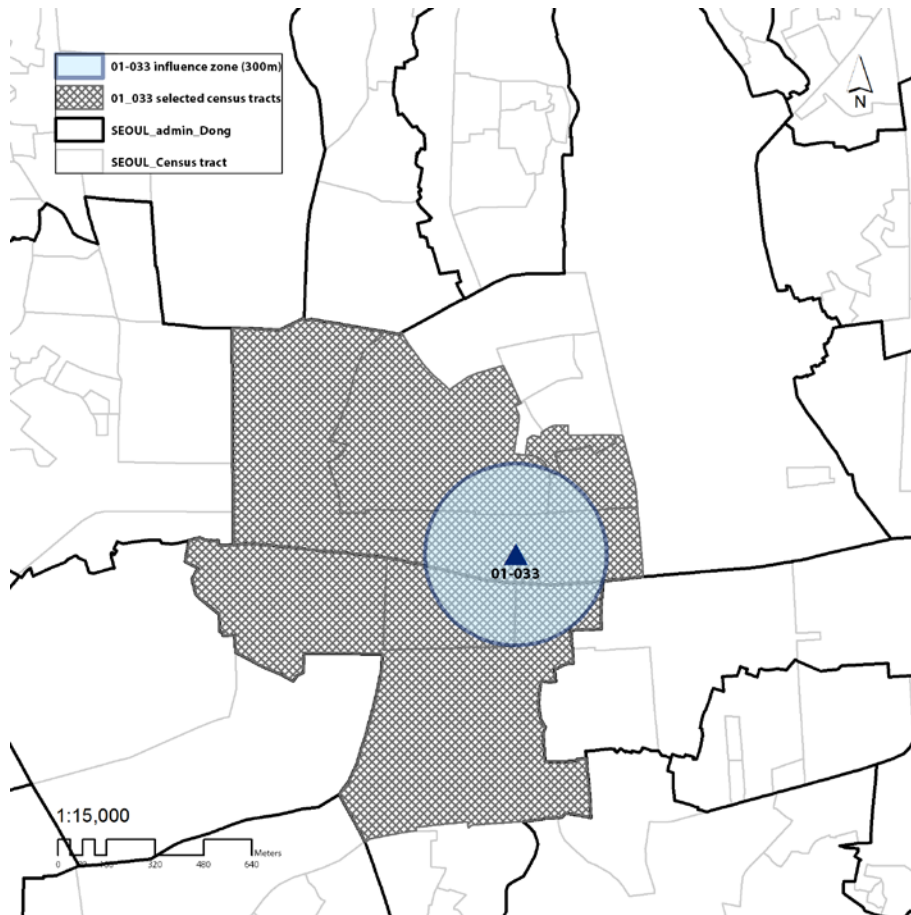Figure 2. Variables lists and model overview.

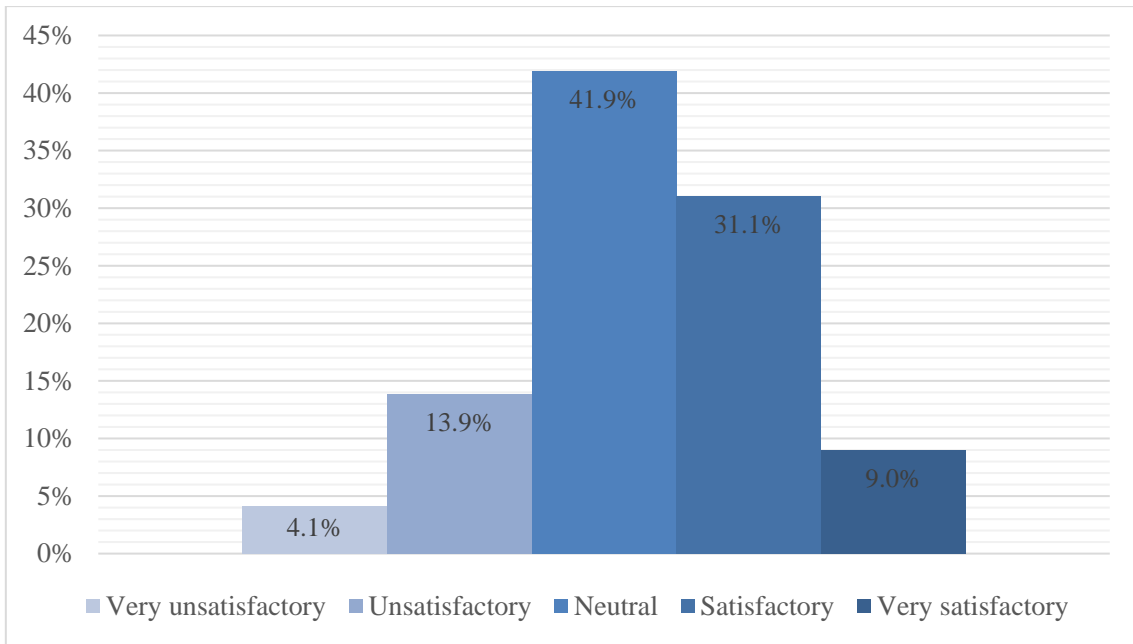Figure 3. Example (01-033 spot) of selected census tracts within a buffer zone (300m).

Figure 4: Pedestrian satisfaction distribution.

Figure 5. Logistic regression coefficient analysis for the visiting frequency of pedestrians (box = insignificance estimates (sig<0.1), first time = reference).

Figure 6. Logistic regression coefficient analysis for the travel purpose. (box = insignificance estimates (sig<0.1), "home" = reference). The categories are ordered into the Utilitarian and Personal groups for convenience.

Figure 7. Logistic regression coefficient analysis for some of the numeric land-use variables. Part (a) shows coefficient estimates for land-use-categorized building counts, while part (b) shows them for land-use-categorized building floor area variables.

Figure 8. Logistic regression coefficient estimates for the effect of residence district. Part (a) shows the estimates for a negative walkability experience, part (b) for the positive walkability experience.

Figure 1. Seoul Pedestrian Interview Spots (total: 1,170).

Figure 2. Variables lists and model overview.

Figure 3. Example (01-033 spot) of selected census tracts within a buffer zone (300m).
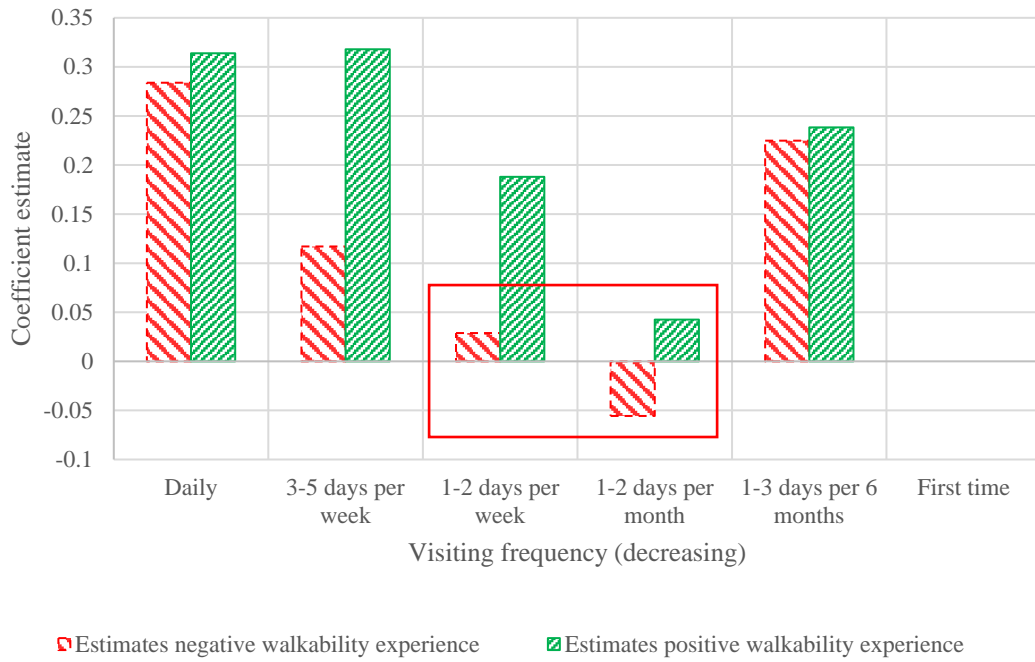
Figure 4: Pedestrian satisfaction distribution.

Figure 5. Logistic regression coefficient analysis for the visiting frequency of pedestrians (box = insignificance estimates (sig<0.1), first time = reference).
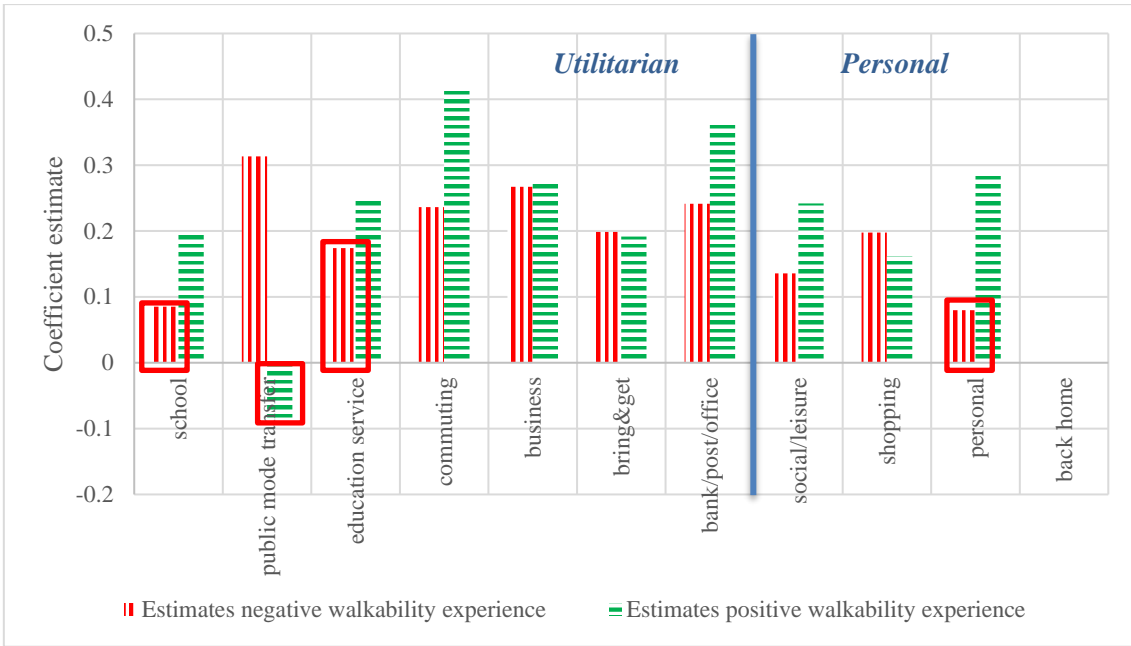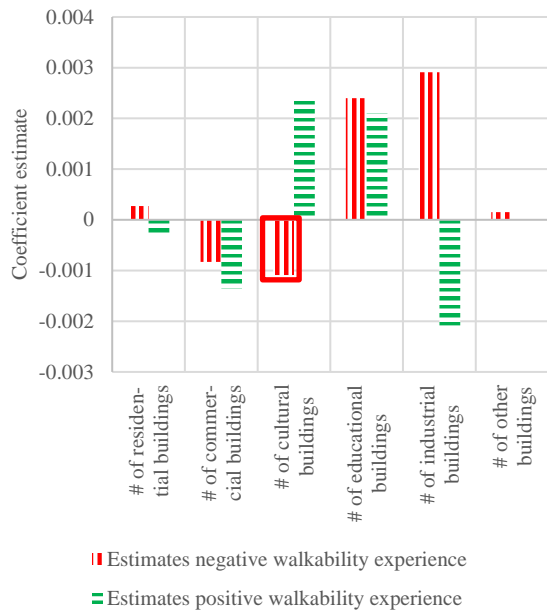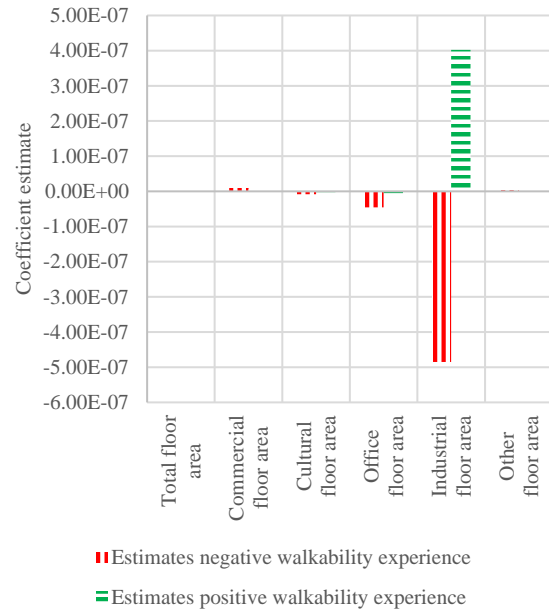
Figure 6. Logistic regression coefficient analysis for the travel purpose. (box = insignificance estimates (sig<0.1), "home" = reference). The categories are ordered into the *Utilitarian* and *Personal* groups for convenience.
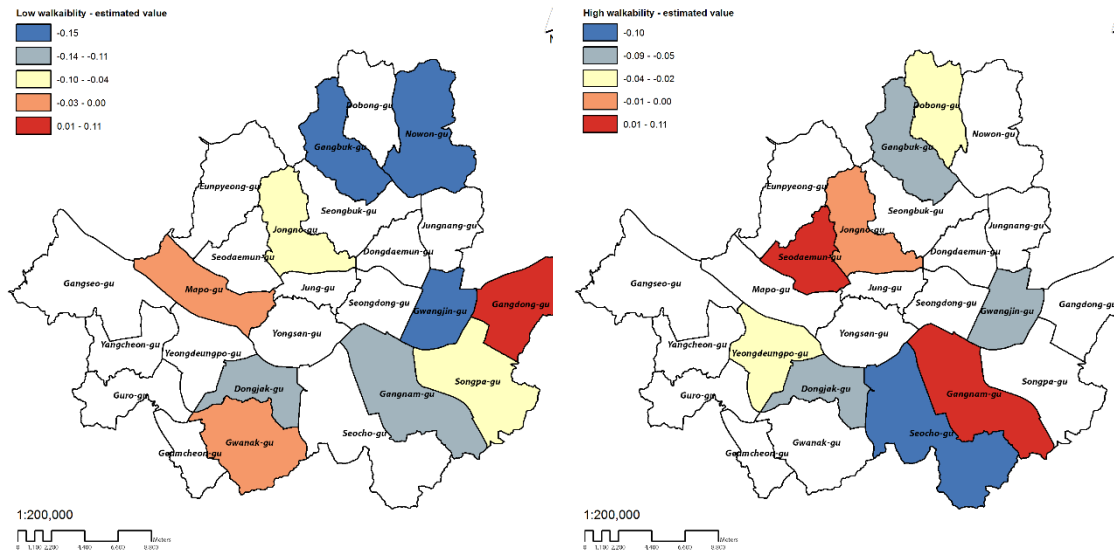
(a)

(b)

Figure 7. Logistic regression coefficient analysis for some of the numeric land-use variables. Part (a) shows coefficient estimates for land-use-categorized building counts, while part (b) shows them for land-use-categorized building floor area variables[1].

---

[1] The box encloses coefficient estimates that were insignificant at the 90% significance level.

(a)   (b)

Figure 8. Logistic regression coefficient estimates for the effect of residence district. Part (a) shows the estimates for a negative walkability experience, part (b) for the positive walkability experience[2].

---

[2] Only coefficient estimates significant at a significance level of 5% are shown. Mind that the colour ranges are different in both legends. The reference category in the logistic regression was "Others", i.e. having a residence outside of the 25 Seoul districts.

Appendix

Group-variables lists and definition

TYPES	VARIABLES	DEFINITION & CODES	Total Mean (S.D.)
Dependent Variable	Walkability	How comfortable to walking this place? (1. Negative; 2. Neutral; 3. Positive)	2.22 (0.73)
Personal	P_GENDER*	Male (0) or Female (1)	0.55 (0.5)
	P_SURVEY_TIME	1. AM 8-11; 2. AM 11- PM 2; 3. PM 2-5; 4. PM 5-8	-
	P_SURVEY_DATE*		-
	P_SURVEY_DAY*	1. Tuesday; 2. Wednesday; 3. Thursday; 4. Friday	-
	P_AGES*	1. 15-19; 2. 20-29; 3. 30-39; 4. 40-49; 5-50-59; 6. Over 60	-
	P_RESIDENCE_CODE	11. Seoul; 99. Other place	-
	P_SEOUL_SUB*	1-25: specific administrative districts (Gu) in Seoul; 99: Other place	-
	P_REGION*	0. Similar places (survey spots in the residential area) 1. Stranger (mismatched with residential area)	0.55 (0.5)
	P_VISIT_PURPOSE*	1. Commuting; 2. Business; 3. School; 4. Education service; 5. Personal; 6. Shopping; 7. Bank/post/office; 8. Bring get; 9. Social & leisure; 10. Public transit mode transfer; 11. Back home	-
	P_VISIT_FREQ*	1. Every day; 2. 3-5 days per week; 3. 1-2 days per week; 4. 1-2 days per month; 5. 1-3 days per 6 month; 6. First visit	-
	P_VISIT_PARTY*	1. Alone; 2. With family or relative; 3. With friend or colleagues; 4. Etc.	-
	P_PRE_USED_MODE*	1. Walking; 2. Public transit; 3. Automobile; 4. Taxi; 5. Auto-bicycle/bike; 6. Etc.	-
	P_JOB*	1. Professional & free-lancer; 2. Officer & engineer; 3. Business & management; 4/ Sales & service; 5. Temporary & physical labor; 6. Product & logistics; 7. Housewife; 8. Students; 9. Self-employed; 10. Unemployed &Etc.	-
Facilities	F_WIDTH_WALK_ROAD*	Walking road width (m)	4.18 (2.3)
	F_CENTRAL*	Presence of central lines (0. No; 1. Yes)	0.61 (0.49)
	F_NUM_LANES*	Number of total road lanes (0. No; 1. Yes)	3.47 (2.64)
	F_BUS_LANE*	Presence of central-bus only lanes (0. No; 1. Yes)	0.11 (0.31)
	F_ OBSTACLE*	Presence of walking obstacles (0. No; 1. Yes)	0.94 (0.23)
	F_BRAILLE_BLOCK*	Presence of braille block (0. No; 1. Yes)	0.32 (0.46)
	F_WALK_ROAD*	0. Mixed traffic street; 1. Pedestrian only; 3. Pedestrian with bicycle	0.75 (0.54)
	F_FENCE*	Presence of sidewalk fence (0. No; 1. Yes)	0.19 (0.39)
	F_SLOPE*	Presence of slope (0. No; 1. Yes)	0.25 (0.43)
	F_CROSS*	Presence of crossings (0. No; 1. Yes)	0.57 (0.5)
	F_BUS_STOPS*	Presence of bus transit station (0. No; 1. Yes)	0.33 (0.47)
	F_SUBWAY*	Presence of subway station gates Presence of bus transit station (0. No; 1. Yes)	0.14 (0.35)
Pedestrian Volume	V_TUE_THR	Pedestrian Volume Survey in Tuesday or Thursday (Cannot Recognize)	4546.01 (4725.36)
	V_WED	Pedestrian Volume Survey in Wednesday	4672.09 (5032.85)
	V_FRI	Pedestrian Volume Survey in Friday	4504.7 (4571.12)
	V_SAT	Pedestrian Volume Survey in Saturday	4549.37 (4715.49)
	V_AVG_WEEKDAY	Pedestrian Avg. Volume of Weekdays	4120.36 (5649.67)
	V_AVG_WEEKEND	Pedestrian Avg. Volume of Weekends	500.56 (652.33)
	V_07_08*	Pedestrian Volume During 7 To 8	4463.61 (4832.37)

	V_08_09	Pedestrian Volume During 8 To 9	891.15 (1013.92)
	V_09_10*	Pedestrian Volume During 9 To 10	675.09 (688.26)
	V_10_11	Pedestrian Volume During 10 To 11	623.55 (618.29)
	V_11_12	Pedestrian Volume During 11 To 12	715.71 (727.08)
	V_12_13*	Pedestrian Volume During 12 To 13	957.35 (1000.55)
	V_13_14	Pedestrian Volume During 13 To 14	928.22 (1041.12)
	V_14-15*	Pedestrian Volume During 14 To 15	889.28 (972.62)
	V_15_16	Pedestrian Volume During 15 To 16	979.23 (1046.96)
	V_16_17	Pedestrian Volume During 16 To 17	1016.21 (1109.48)
	V_17_18	Pedestrian Volume During 17 To 18	1072.98 (1214.11)
	V_18_19	Pedestrian Volume During 18 To 19	1471.48 (1624.08)
	V_19_20	Pedestrian Volume During 19 To 20	1209.94 (1496.98)
	V_20_21	Pedestrian Volume During 20 To 21	977.66 (1276.82)
	V_TIME_ZONE_1	Pedestrian Volume During 8 to 11	2189.79 (2199.52)
	V_TIME_ZONE_2	Pedestrian Volume During 11 to14	2601.27 (2711.17)
	V_TIME_ZONE_3	Pedestrian Volume During 14 to17	2884.72 (3107.23)
	V_TIME_ZONE_4	Pedestrian Volume During 17 to 20	4732.05 (5512.99)
	actual_timeZone_vol		3290.04 (4805)
	actual_DOW_vol		4689.76 (5085.81)
	Crowdedness*		0.993 (0.374)
Land-Use	L_T_POP*	The Number of Population (EA)	20.09 (8.26)
	L_T_HH*	The Number of Households (EA)	11658.47 (5054.42)
	L_T_COMP*	The Number of Companies (EA)	4284.72 (1831.63)
	L_T_EMP*	The Number of Employees (EA)	1054.77 (1619.91)
	L_T_HOUSING*	The Number of Housing (EA)	1469.45 (901.87)
	L_B_TOTAL_N	The Number of Total Building (EA)	3894.85 (5318.02)
	L_B_RES_N*	The Number of Residential Building (EA)	3009.14 (1410.83)
	L_B_COMMER_N*	The Number of Commercial Building (EA)	754.05 (492.93)
	L_B_CUL_N*	The Number of Cultural Building (EA)	26.43 (47.98)
	L_B_EDU_N*	The Number of Educational Building (EA)	19.86 (24.54)
	L_B_OFFICE_N	The Number of Bank/Office/Public Building (EA)	24.74 (32.74)
	L_B_INDUS_N*	The Number of Industrial Building (EA)	37.32 (39.14)
	L_B_ETC_N*	The Number of Other Building (EA)	7.6 (65.23)
	L_B_TOTAL_A*	The Area of Total Building (km²)	568.25 (704.7)
	L_B_RES_A	The Area of Residential Building (km²)	11.93 (50.61)
	L_B_COMMER_A*	The Area of Commercial Building (km²)	2.24 (10.63)
	L_B_CUL_A*	The Area of Cultural Building (km²)	0.56 (7.44)
	L_B_EDU_A	The Area of Educational Building (km²)	0.63 (4.31)
	L_B_OFFICE_A*	The Area of Bank/Office/Public Building (km²)	3.18 (39.69)
	L_B_INDUS_A*	The Area of Industrial Building (km²)	0.48 (2.09)
	L_B_ETC_A*	The Area of Other Building (km²)	0.05 (0.39)

	CENSUS_N	The Number of Census Tracts (EA) within 300m ranges from specific survey spots)	4.79 (20.57)
	CENSUS_AREA	The Area of Total Census Tract (㎢) within 300m ranges from specific survey spots)	1 (0.88)
	ENTROPY*	Land-Use Mix (0-1)	0.93 (0.03)
Weather	W_MAX_HUMID	The maximum humidity (Φ) of the specific day in Seoul	88.49 (9.6)
	W_MAX_TEMP	The maximum temperature (℃) of the specific day in Seoul	18.85 (7.74)
	W_EVENTS*	1. Sunny; 2. Cloudy; 3. Rainy; 4. Snow	-

Note*: Retained in the model with all variable themes