IOP Conf. Series: Materials Science and Engineering 466 (2018) 012012 doi:10.1088/1757-899X/466/1/012012

A Study on Outlier Detection and Its Effects: Using the South Korean Residential Condition Survey

Jungmin Choi

Department of Architecture, Konkuk University, Faculty of Architecture, room A610, 120 Neungdong-ro, Gwangjin-gu, Seoul 143-701, South Korea
Email: jmchoi@konkuk.ac.kr

Abstract. This study analyzed outliers which give a significant impact on reliability in the current Big Data era. We used the Housing Condition Survey (HCS) by the Ministry of Land, Infrastructure and Transport in Korea as a main dataset. We focused on the residential satisfaction of the respondents, detected the outliers, and performed an empirical analysis for impact factors. Although we used the Mahalanobis distance method, which is a common method for outlier detection, we found that this outlier detection method was unreliable. Alternatively, the Naive Bayes Classification method was utilized to detect the outliers and to verify the impact factors. This choice of method was based on the fact that the high correlation among the demographic characteristics and residential satisfaction of respondents are critical elements in the Naive Bayes Classification. The findings include that firstly, about 2,400 samples (12% of total) of the '2014 Housing Condition Survey' were detected as outliers. Secondly, it was observed that the tendency of positive over-estimation about questions from the residential satisfaction of respondents in HCS is common. Thirdly, in order to reduce the occurrence of outliers in HCS, it is necessary to lessen the stress of respondents by avoiding long questions in table form.

1. Introduction

Recently, studies using big data have been actively conducted in diverse fields. Taking advantage of the informational stream, diverse studies utilizing IT have been attempted in the fields of urban planning and architectural. However, efforts to verify outliers or their reliability from the collected data have been made less frequently. Therefore, this study aims to examine outliers using the residential condition survey of the Ministry of Land, Infrastructure and Transport of South Korea, as well as introducing a process to detect outliers and analysing the disturbing effects of outliers while presenting improvement measures.

2. Theoretical Considerations

2.1. Outliers and Detection Method

The term “outliers” refers to values that are numerically separated from a larger group of values that show general behaviour among collected data. Although outliers may be meaningful data that represent a large change in a phenomenon, outliers due to errors can cause a bias in estimation, leading to the degradation of the quality of large-scale statistical surveys; their effects are even larger in the case of periodic surveys. Methods to detect outliers include statistical tests, depth-based approaches, deviation-based approaches, distance-based approaches, density-based approaches, and high-dimension-based approaches. In addition, diverse new methods have been presented and are evolving steadily.

2.2. South Korean Residential Condition Survey and Satisfaction Evaluation Items

The periodic residential condition surveys by the Ministry of Land, Infrastructure and Transport of South Korea have been carried out every two years since 2006. Their purpose is to investigate the general national housing and residential environments and to establish housing policies that coincide with the features of diverse classes. The collected survey data are representative survey data utilized by the South Korean government to grasp basic statistics for housing supply and demand analysis, the trends of households below the minimum housing standard, housing consumption level, residential environment satisfaction, and housing values when establishing comprehensive housing plans. These surveys systematically extract about 20,000 to 30,000 respondents nationwide; their information helps to identify trends of residential conditions through about 300 housing-related questionnaire items.

This study was conducted focusing on outlier detection in "residential environment satisfaction evaluation" among the questionnaire items of the residential condition surveys. Table 1 summarizes contents related to residential environment satisfaction in the residential condition surveys. The residential environment satisfaction evaluation items in the residential condition surveys are mainly composed of those centered on "convenience" among the five ideologies of residential environments. In detail, they consist of 12 to 16 individual items; at the end of the evaluation items there are items to evaluate the overall satisfaction with houses and overall residential environments.

Table 1. Composition of evaluation items on residential environment in South Korean periodic residential condition survey in 2006-2014

		Year 2006	Year 2008	Year 2010	Year 2012	Year 2014
Survey supervision		KRIHS	KRIHS	KRIHS	Land & Housing Institute	KRIHS
No. of samples		30,201	30,156	33,000	33,000	20,205
Total questionnaire items		350	299	280	257	247
Items related to residential satisfaction		12	15	16	12	15
Likert scale*		4 Point	4 Point	4 Point	5 Point	4 Point
Common evaluation items on residential environment	q.1	Access easiness to Market, Department store, etc.				
	q.2	Access easiness to Hospital, Medical welfare facilities, etc.				
	q.3	Access easiness to City hall, Local authority's office, Police, etc. Public institute				
	q.4	Access easiness to Cultural facilities, Amusement park, Green park, Waterfront, etc.				
	q.5	Access easiness to Bus, Subway, etc. Public transportation				
	q.6	Convenience to Parking facilities				
	q.7	Educational environment to those facilities for Preschool child, Middle or High School, Private educational facilities				
	q.8	Status of Crime prevention for security				
	q.9	Status of Cleaning and waste disposal				
	q.10	Status of Relationship with neighborhood				
	q.11	Overall status of residential environment				

2.3. Residential Environment and Residential Satisfaction Influencing Factors

The residential environment is defined as the whole of the living environments that surround the residential place. In a limited sense, it includes the environment around the house; in a broader sense, it includes the social, economic, and cultural environments. In addition to safety, health, convenience,

and comfort, the contents of the residential environment include sustainability as well as considerations about both the present and the future (Asami et al., 2003). The residential environment as such greatly affects the formation of personality as a base for living as well as a container of human life.

Residential satisfaction means satisfaction with not only the physical properties of the house but also the overall residential environment, including the physical and social environments around the house. It is made up of subjective and emotional positive or negative responses based on the resident's taste, preference, and experience. Since residential satisfaction depends on subjective evaluation, the evaluation changes according to conditions and levels of satisfaction are distributed, centering on the middle rather than extremes (Kim et al., 2011). Therefore, there are risks of causing damage to the original representativeness and risks of errors.

3. Outliers and Study Model Setting

3.1. Basic Analysis of Outliers in Satisfaction Evaluation in the Residential Condition Survey

Since "satisfaction" is measured by individuals' subjective judgment and empirical "value" and does not require answers about objective facts, the probability of outliers resulting from the respondents' unfaithfulness may be doubled due to the "fatigue effect" of consecutive questionnaire items on respondents. In other words, when respondents must answer several subjective questionnaire items in a sequence, they develop strong tendencies to choose median values or general values rather than extreme values. In the case of consecutive evaluation items based on the same evaluation scale in the form of tables, the possibility of subjects responding simply and repeatedly or arbitrarily responding in mechanical ways will become higher, and they are less likely to have read the questionnaire items properly.

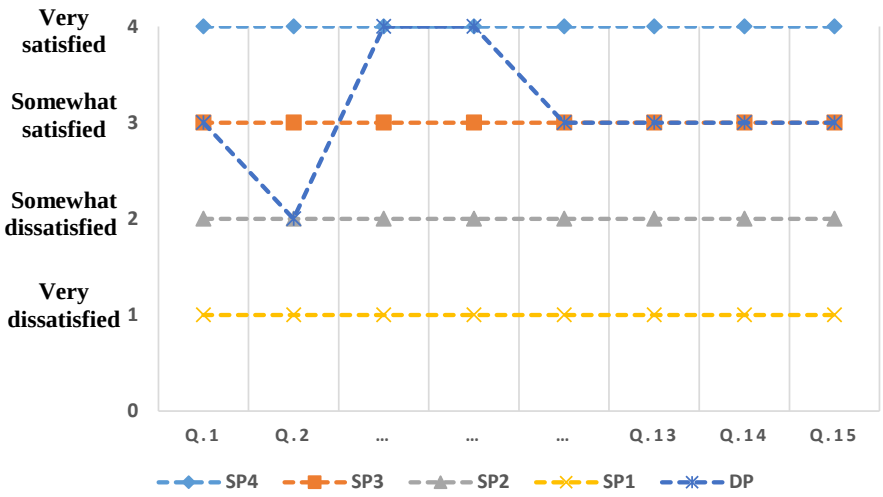

Figure 1. Diagram of the same patterns as suspected outliers

Bearing these factors in mind, in this study the response patterns of those who answered with the same response values (e.g., repetition of consecutive identical values such as 333…333, 444…444) among the entire approximately 140,000 residential condition survey respondents in 2006-2014 were set as candidates for outliers, and basic analyses were conducted. Figure 1 shows the same response patterns in the residential satisfaction evaluation in the periodic residential condition surveys where, for example, "SP.1" means responses with "1 point (very dissatisfied)" in the Likert evaluation scale for the same patterns. That is, SP.1 ~ SP.4 indicate the patterns in which individual respondents responded with the same scores in the residential environment satisfaction evaluation items. These response patterns are judged to be sufficient for suspecting that a response is an outlier.

Table 2 shows the same response patterns aggregated by year. The ratio of the same response patterns suspected as outliers to the entire response patterns was shown to be 9.24% on average. In other words, nearly 10% of all respondents responded with the same scores to the satisfaction evaluation item. In particular, in the 2014 survey, the share of the same response patterns (SP) increased to 15.7%, which is about 1.7 times the 5-year average, indicating that the ratio of outlier suspicion is high, corresponding to 3,171 out of 20,205 respondents in total.

Also, most of the response patterns are SP.3 ("generally satisfied"), indicating that the responses are quite biased toward "satisfied" as the relevant respondents of about 8% of all respondents. On reviewing the response patterns concretely based on the 2014 data, the tendency biased toward "satisfactory" is clearly revealed in the residential environment satisfaction evaluation, as the ratio of responses indicating dissatisfaction (integration of "very dissatisfied" and "dissatisfied") is 0.25% of the entire responses, while the ratio of responses indicating satisfaction (integration of "generally satisfied" and "very satisfied") is 15.5%, accounting for 98% (15.46/15.71) of the same patterns suspected as outliers. In addition, in the residential condition survey in 2014, the same patterns suspected as outliers appeared much more frequently, to reach almost two times the frequency seen in the past four surveys, in which the ratios were 7.04~9.92% (average 7.62).

Table 2. Ratio of the same patterns as suspected outliers

Year	Classification	SP.1	SP.2	SP.3	SP.4	Sum (N/%)
		Very dissatisfied	Dissatisfied	Somewhat satisfied	Very satisfied	
2006	Frequency	19	196	2,520	260	2,995
	Ratio	0.06	0.65	8.34	0.86	9.92
2008	Frequency	17	58	1,993	147	2,216
	Ratio	0.06	0.20	6.61	0.49	7.35
2010	Frequency	17	82	2,092	133	2,324
	Ratio	0.05	0.25	6.34	0.40	7.04
2012*	Frequency	30	28	1,376	296	2,046
	Ratio	0.09	0.08	4.17	0.90	6.20
2014	Frequency	6	44	2,724	397	3,171
	Ratio	0.03	0.22	13.5	1.96	15.71
Average ratio		0.06	0.24	7.79	0.92	9.24

* Ratio of "Somewhat satisfied" was considered in 2012

3.2. Application of Existing Outlier Detection Methods and Bayesian Analysis

Among outlier detection methods, "Mahalanobis distance," which is generally used widely, was applied to the residential condition survey in 2014 to extract outliers. Since Mahalanobis distances can be calculated only when quantitative continuous variables are used, three variables were used in this residential condition survey to extract outliers based on Mahalanobis distances: income, age, and housing satisfaction, which are relatively closely correlated with the satisfaction evaluation. As a result, the outlier boundary point distance of Mahalanobis distances was calculated as 8.99955, and out of a total of 20,039 samples (excluding missing values), the number of samples that were outside the distance and thus were extracted as outliers was shown to be 1,253 (6.25% of the entire sample). On reviewing the responses of these 1,253 respondents regarding residential environment satisfaction, it could be seen that most of them responded as being "satisfied" (3 points) and "very satisfied" (4 points). However, when outlier detection methods such as Mahalanobis distance are used, although a group of candidates suspected as outliers can be extracted, the grounds are insufficient to concretely confirm outliers.

Therefore, in this study, the fact that a certain correlation is established between residential environment satisfaction and the features of respondents was noted and outlier extraction reflecting the features of respondents was attempted. Table 3 shows the relationship between satisfaction and the features of respondents. As can be seen in the table, women are more satisfied than men, and older people are more satisfied than younger people. In addition, those with higher education levels, those who are healthier, those with stronger family relations, and those with higher quality and more convenient residences are more satisfied (happier). Therefore, if the features of the respondents can be reflected in connection with the evaluation of the residential environment satisfaction, the detection of outlier candidate groups will be easy and highly credible grounds to finalize the groups can be presented after extracting candidates for outliers. Taking notice of the foregoing, in this study, a decision was made to apply Bayesian statistics.

Table 3. Relationship between satisfaction and various influencing factors

Influencing factors		Subject matter
Demographics	Gender	Female than male ↑
	Age	Young people(29~29 age) ↓, more than 60s↑
	Education	The higher the education level ↑
	Income	The higher the income ↑
	Martial status	Married than single ↑
	Parenting	If there is a child, life happiness is high ↑, Marriage happiness ↓
	Family	The stronger the family relationship ↑
	Health	The healthier ↑
	Sleeping hours	The longer sleeping hours ↑
	Religion	The higher the faith ↑
	Finance	The happiness of empirical consumption is higher than the material consumption ↑
Lifestyle, Behavior	Smoking	Happiness in smoking ↓
	Drinking	Happiness in drinking ↑
	Lottery ticket	The higher the purchase rate, the higher the happiness ↓
	Vote	Happiness is high in voting ↑
Housing, Residential environment	Housing	The higher the quality of housing ↑(residential satisfaction)
	Convince of living	The higher the convenience of living ↑
	Natural environment	The better the natural environment ↑
Socio-economic characteristics	Employment	Unemployment experience and anxiety, retirement ↓
	System	Promoting individual intellectual activities ↑ The more successful social systems ↑ The more important the social welfare ↑
	Community	Political interest, mutual understanding, higher local awareness ↑
	Economy	When it comes to inflation ↓
Psychological characteristics	Personality	The higher the brightness, optimism, gratitude, trust, socialization, and responsibility ↑
	Anxiety	The higher the frequency of anxiety, depression, and loneliness ↓
	Stress	The higher the stress ↓
	Suicide	Happiness ↓ Probability of suicide ↑

* I have summarized the relationship by referring to Fukuda(2010), Sasaki(2013) etc.

3.3. Study Model Setting

3.3.1. Estimation method of the study model

Table 4. Features to be applied to the research model

		Items related to residential environment
Demographic characteristics	Gender	Man/Woman
	Age	20's (0~29), 30's (30~39), 40's (40~49), 50's (50~59), 60's (60~69), Over 70's (70~)
	Education	Below elementary school, junior graduation, high school, university graduation or more
	Spouse	Living together / Non-living
	Children	Living together / Non-living
	Income	Step 10 of the income quintile
	Job	Yes / No
Residential characteristics	Housing Type	Detached house, Apartment, Mansions, etc.
	Occupancy Type	Home ownership, Jeonse, Rent, Free
	Housing Area	Small(0~60m^2), Medium(60~85m^2), Large (85m^2 ~)

The estimation method of the study model is as follows. First, a priority probability is calculated based on information about past features of respondents, using the residential condition survey data for the period of 2006 ~ 2010 (Stage S1). Thereafter, a posterior probability is inferred based on the information about the features of respondents from the 2014 residential condition survey (Stage S2). Finally, the measured values of respondents (satisfaction levels actually responded) and the estimated posterior probability in this case is calculated to detect candidates for outliers (Stage S3).

3.3.2. Extraction of the features of respondents to be applied to the study model

In the Bayesian inference of this study model, conditional features of respondents should be set. Variables (feature variables) that affect residential environment satisfaction were extracted first; variables that are closely correlated with the relevant satisfaction were secondarily extracted through correlation analyses between the variables extracted as such and the satisfaction. The resultant variables are set forth in Table 4. In the Bayesian approach used in this study, which requires conditional probability, the information on the features of respondents corresponds to the foregoing. In Table 4, the features of respondents were indicated after being classified into demographic features (such as sex and age) and residential condition features (such as house type and occupation type).

4. Empirical Analysis with the Study Model

4.1. Detection of Candidates for Outliers Using The Study Model

In this study, 1) respondents (common respondent), with no change in the posterior probability in any learning data when the learning data for Bayesian estimation was applied after being divided into various forms, were first classified into candidates for outliers, and 2) the increments of changes in the number of respondents according to increases in the value of the posterior probability when the value of the posterior probability was increased into a certain section (5%) were examined to set the point where the ratio value of the increments became the smallest as a boundary value. When the logic is followed, if 1 and 2 are applied, the boundary values of outlier candidate detection will be 20 ~ 30%. However, as will be described later, when the averages are compared before and after excluding detected outliers in the evaluation of satisfaction with 11 kinds of residential environments, the amounts of changes in 2 are larger in cases where the boundary value is 30%. Therefore, those values were set as boundary values.

Table 5. Bayesian estimation corresponding to the measured value Outlier detection by posterior probability

	5%	10%	15%	20%	25%	30%	35%	40%
t06	16	159	619	1456	2812	4763	6922	9257
t08	20	166	594	1186	2050	3310	5127	7430
t10	59	265	621	1134	1900	2963	4305	6187
t06-10	31	196	611	1159	1985	3101	4693	7007
Common respondent*	11	130	440	957	1639	2570	3829	5584
Increase**	-	119	310	517	682	931	1259	1755

* Common respondent: When the year 2006, 2008, 2010, 2006-2010 are used as the learning data, the interval value of the extracted posterior probability indicates a common respondent.
** Increase: Represents the increment of the value of the posterior section relative to the value of the preceding section of the posterior probability section.

4.2. Analysis of the Effects of Outliers On Satisfaction

4.2.1. Simulation of changes in satisfaction due to outliers

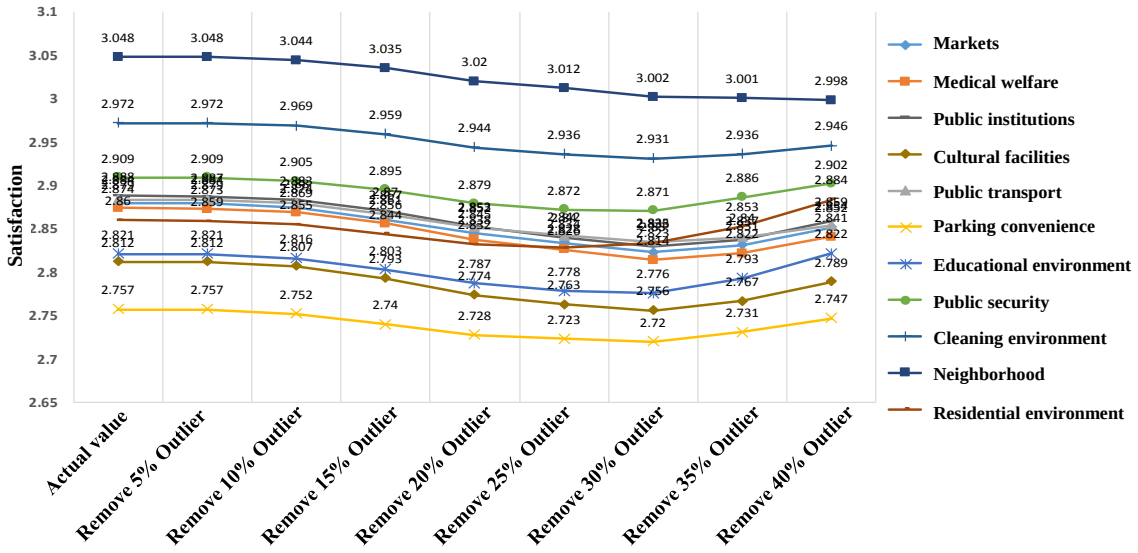

Figure 2. Distribution of the mean value of the satisfaction level when the outlier candidate is removed by the estimated probability interval (Year 2014)

Outlier candidate groups according to the posterior probability obtained through Bayesian estimation and changes in average values from satisfaction evaluation when the outlier candidate groups have been removed will be simulated. Figure 2 shows the distribution of average values of satisfaction by questionnaire item when the posterior probability values estimated through the method explained earlier were divided into 5% ranges and the outlier candidates were finalized as outliers by the relevant range of outlier candidate groups and were removed. In the figure, the leftmost row shows “actually measured values,” which are the average values of satisfaction by questionnaire item in the 2014 periodic residential condition survey.

On reviewing changes in the average values of satisfaction after removing the outliers of the relevant posterior probabilities in all questionnaire items, it is interesting to note that great changes occur when the posterior probability is around 25 ~ 30%. As can be seen in Figure 2, the largest gaps between estimated values and measured values occur when the posterior probability has been set as 30% and the outlier candidate groups have been removed. This is consistent with the results set forth in Table 5 and can be said to prove that setting the estimated posterior probability to detect outliers in this study model as 30% is valid.

4.2.2. Comparison and analysis of satisfaction evaluation of those that are suspected as outliers
Table 6 shows the same response patterns of respondents extracted based on the estimated posterior probability of 30%. In the table, respondents with the same pattern (SP.1 to SP.4) account for 23.3% of 599 out of 2,570 outlier suspects. In the table, those respondents who answered with the same patterns (SP.1 to SP.4) account for 23.3% or 599 out of the entire 2,570 respondents suspected as outlier. In particular, the share of the same response patterns of outlier respondents extracted based on the estimated posterior probability patterns (SP) attracts attention. That is, when the shares are calculated again based on the same response patterns of the outlier respondents extracted based on the estimated posterior probability (set to 30%), the share of SP.4 becomes 90.2% (540/599), that of SP.2 becomes 8.1 % (49/599), that of SP.1 becomes 1.7% (10/599), and that of SP.3 becomes 0% (0/599). This is a very encouraging result. The response SP.4 accounted for 2.82% with 570 respondents (570/20,205) in the measured values. However, among outliers, the response SP.4 accounted for 21.0% with 540 out of the 2,570 respondents suspected as outliers and 90.2% (540/599) of respondents with the same response patterns, indicating that most outliers with the same response patterns are in the form of SP.4.

Table 6. Frequency of respondents with the same pattern for each posterior probability and ratio to aggregated units

			Bayesian inference posterior probability								2014 Actual value
			5%	10%	15%	20%	25%	30%	35%	40%	
The same patterns as suspected outliers	SP.1	Frequency	2	8	9	9	9	10	12	13	14
		Ratio	18.18	6.15	2.05	0.94	0.55	0.39	0.31	0.23	0.07
	SP.2	Frequency	-	-	2	9	34	49	59	66	67
		Ratio			0.46	0.94	2.07	1.91	1.54	1.18	0.33
	SP.3	Frequency	-	-	-	-	-	-	-	4	3,513
		Ratio								0.072	17.39
	SP.4	Frequency	8	96	262	415	484	540	556	570	570
		Ratio	72.73	73.85	59.55	43.37	29.53	21.01	14.52	10.21	2.82
Sum of same patterns (SP.1~SP.4)			10	104	273	433	527	599	627	653	4,164
Common respondent			11	130	440	957	1,639	2,570	3,582	5,115	20,205

This result is interpreted to be caused by the fact that this study model takes the posterior probability of the Bayesian approach, reflecting diverse characteristics of the respondents, and as a result many of those who evaluate satisfaction extremely despite insufficient grounds (in particular, SP.4: very satisfied) were extracted. That is, although not seen in measured values because they were covered by values such as SP.3 (generally satisfactory), those respondents with no certain regulation between their features and satisfaction evaluation were extracted as suspected outliers in the process of inference reflecting the features of respondents in this study model. These respondents suspected as outliers exhibit response characteristics extremely or excessively biased toward overestimated satisfaction, as clearly shown by SP.4 (very satisfied).

4.3. Comparative Analysis of Satisfaction Items with and Without Disturbance by Outliers

Figure 3 shows that 1) the average values of satisfaction of all questionnaire respondents and 2) the average values of satisfaction of the questionnaire respondents after removing the 2,570 respondents (Table 6) extracted as suspected outliers according to the study model described earlier by questionnaire item for the 11 common residential environment satisfaction items in the 2014 residential satisfaction survey in comparison with each other. First, as can be seen from the figure, the average values of satisfaction decreased in all questionnaire items after removing those respondents that were extracted as suspected outliers. As explained earlier, this is attributable to the fact that most of the suspected outliers who were extracted in this study have a tendency toward overestimation of satisfaction as SP.4.

Differences in satisfaction scores before and after removing suspected outliers appear by questionnaire item. In Figure 3, the values shown between the solid line and the broken line are the differences (1-2) between the solid line (1, averages of satisfaction of all survey respondents) and the broken line (2, averages of satisfaction of survey respondents after removing suspected outliers) calculated by questionnaire item.

As for the sizes of differences by satisfaction item, the order is as follows: q.2 (easiness to access medical welfare) > q.3 (easiness to access public institutions) > q.1 (easiness to access markets) = q.4 (easiness to access cultural facilities). These are items that evaluate the easiness to access most facilities, and suspected outliers were shown to relatively overestimate these items. In contrast, the tendency toward overestimation seems to be weaker in the case of the following questionnaire items: q.11 (comprehensive residential environment satisfaction), q.6 (parking convenience), and q.8 (public security issues).

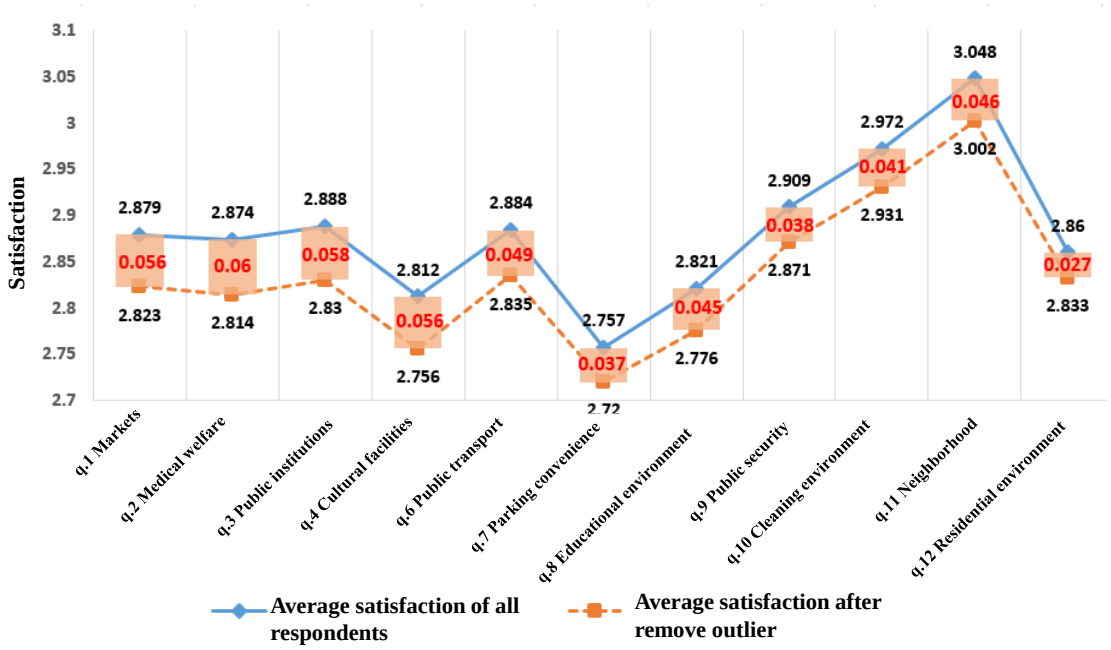

Figure 3. Change of Residential Satisfaction after Removing the Suspected outliers (Posterior probability 30% was applied)

5. Conclusion

Based on the results of detection of outliers and analysis of the disturbing effects of the outliers conducted in a series of processes described above, the following improvements are judged to be necessary.

First, efforts should be made to reduce the questionnaire burden of respondents as much as possible in the case of items that rely on individuals' subjective values, such as residential environment

satisfaction evaluation items. Consecutive questionnaire items to be evaluated by Likert 4-point scales listed in the form of tables make the respondent feel fatigue, which leads to unfaithful responses and results in a high possibility of a considerable increase in outliers. This tendency will be intensified as the number of consecutive questionnaire items increases.

Second, most respondents suspected as outliers extracted with the study model tend to overestimate satisfaction in the same response patterns, especially in the form of SP.4 (very satisfied). Therefore, those respondents who overestimate satisfaction positively with the same value (e.g., SP.4) when a large number of questionnaire items are given consecutively in satisfaction evaluation surveys can be safely classified as suspected outliers. Researchers should first extract them and compare the demographic characteristics and related attribute information to remove respondents with lower causal relationships and recount satisfaction evaluation data.

Third, some respondents who are assumed to be outliers because of their unfaithful responses in residential environment satisfaction evaluation are assumed to show strong tendencies toward unfaithful responses in other evaluations too, such as questionnaire items. Table 7 shows the results of evaluation of Question 23 (individual residential environment satisfaction items) and Question 24 (comprehensive satisfaction with house and residential environments) connected with the results of evaluation of Question 22 (housing satisfaction), indicating which response patterns appear. As can be seen in Table 7, as with the residential environment satisfaction evaluation intensively analysed in this study, the same response patterns are present in the evaluation of housing satisfaction. That is, among the 20,205 respondents in the 2014 residential condition survey, 1,429 respondents (7.1% of all respondents) gave the same responses such as P01 (3,3,3 ... 3,3,3) for 23 questionnaire items, so that the suspected outliers are assumed to have responded unfaithfully not only for certain questionnaire items, but also other items.

Table 7. Satisfaction Patterns on Housing and Residential Environment by Rank

Rank of pattern	Respondents' satisfaction pattern		Frequency	Ratio
	Satisfaction of housing	Satisfaction of residential environment		
P01	3,3,3,3,3,3,3,3	3,3,3,3,3,3,3,3,3,3,3,3,3,3,3	1429	7.1
P02	4,4,4,4,4,4,4,4	4,4,4,4,4,4,4,4,4,4,4,4,4,4,4	282	1.4
P03	3,3,3,3,3,2,3,3	3,3,3,3,3,3,3,3,3,3,3,3,3,3,3	282	1.4
P04	4,4,4,4,4,4,4,4	3,3,3,3,3,3,3,3,3,3,3,3,3,3,3	195	1.0
P05	3,3,3,3,3,3,3,3	3,3,3,3,3,2,3,3,3,3,3,3,3,3,3	141	0.7
P06	4,4,4,4,4,4,4,4	3,3,3,3,3,3,3,3,3,3,3,3,3,4,3	65	0.3
P07	3,3,3,3,3,2,3,3	3,3,3,3,3,2,3,3,3,3,3,3,3,3,3	60	0.3
P08	4,4,4,4,4,3,4,4	3,3,3,3,3,3,3,3,3,3,3,3,3,3,3	55	0.3
P09	4,4,4,4,4,3,4,4	4,4,4,4,4,4,4,4,4,4,4,4,4,4,4	50	0.3
P10	3,3,3,3,3,3,3,3	3,3,3,3,3,3,3,3,3,2,3,3,3,3,3	44	0.2

	SP.1	SP.2	SP.3	SP.4
Frequency	3	20	1429	282
Ratio	0.015	0.099	7.073	1.396

Fourth, to reduce those respondents who are suspected as outliers due to their unfaithfulness, as mentioned above, along with the effort to reduce the burden of questionnaires, questionnaire items should be appropriately divided and distributed at various places on the questionnaire sheet so that respondents are not as conscious of the questionnaire items. Alternatively, the occurrence of the same response patterns can be reduced by appropriately inserting other evaluation scales (e.g., yes or no) in

the middle of the questionnaire. Another way is to integrate items with high correlations frequently found in the current residential environment satisfaction evaluation through cluster analysis and other means, thereby reducing the number of questionnaire items and leading to reduction in the mental burden of respondents, which will eventually reduce the probability of occurrence of outliers.

Fifth, the probability of occurrence of outliers can be reduced by designating concrete spatial ranges of satisfaction in the development of residential environment satisfaction questionnaire items, because the spatial ranges recognized and understood by individual respondents are different from each other. Errors in the recognition of satisfaction evaluation may be caused by questionnaires. As in cases where the foregoing fact has been reflected, for instance, the evaluation of satisfaction in the Korean General Social Survey (KGSS) or the American House Survey (AHS) in the United States presents concrete distances, or the time taken to move, to minimize errors due to respondents' recognition.

As a future task, other analysis methods than the method proposed in this study should be applied to compare and analyze the results and seek further implications.

6. Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2015R1D1A1A01059065)

7. References

[1] Asami Y et al. (2001), *Residential Environment: Methods and Theory for the Evaluation*, University of Tokyo Press.
[2] Kim S et al. (2011), *Personality Traits and Response Styles*, Survey research, 12(2), pp.51-76
[3] Sasaki K (2013), *The Influence of Behaviour and Custom on Subjective Happiness*, Nagoya Gakuin University review, 49(3), pp. 27-42
[4] Toyoda S (2010), *Happiness Degree and Living Economics*, pp.1-10