

# Analysis of Trade Area for Retail Industry Store Using Consumer Purchase Record

Sachiko Iwasaki<sup>1</sup>(✉), Ko Hashimoto<sup>2</sup>, Kohei Otake<sup>1</sup>,  
and Takashi Namatame<sup>1</sup>

<sup>1</sup> Chuo University, Bunkyo-ku, Tokyo, Japan  
sachikoi@namalab.org,

{otake,namatame}@indsys.chuo-u.ac.jp

<sup>2</sup> Graduate School of Chuo University, Bunkyo-ku, Tokyo, Japan  
a12.w6dw@g.chuo-u.ac.jp

**Abstract.** For retail industry such as supermarket and convenience store, it is important to understand customers. In marketing perspective, to match marketing activity to customer needs is one of the most important strategy for retail industry. In this study, we focus on trade area of retail store. If we can grasp the trade size of a store, manager can plan optimal strategy, e.g. how to spend for advertise activity and where we should open a new store. In this study, we use ID-POS data which is the purchase record with customer identification data of a super market chain and calculate the trade area radius, then we show the cause and effect model to estimate trade area size using store causal data. Moreover, we evaluate our model and discuss how to be decided the trade area size.

**Keywords:** Trade area · ID-POS data · Consumer purchase behavior

## 1 Introduction

Consumer purchase behavior is one of the main research areas of marketing. In recent years, many kind of consumer behavior or activity data in the field of marketing can be obtained, i.e. ID-POS data which is purchase record for each customer in a store or customer attributes. Hence, many stores or retail company want to utilize these data for more effective and efficient for marketing activities.

For retail store, it is an important to grasp the trade area, because if the store can grasp his/her trade area, then he/she can advertise efficiently in the trade area or grasp the needs of main trade area. Especially for Japanese retail store, e.g. supermarket and department store, folding flyer in newspaper is very popular advertising tool, thus to grasp the trade area is very important. Moreover, when manager plan to open a new store, to analyze the potential selling intense is very important issue. As shown these topics, the trade area is one of important factor for store managing.

In this study, we analyze trade areas of stores of a supermarket chain in Japan. Based on the purchase record of each store, we calculate a radius of trade area, and analyze some variables which effects on the size of trade area.

## 2 Related Studies and Objective of Our Study

One of the most famous trade area analyzing model is Huff model (Huff 1963). Huff model is a probabilistic choice model based on attraction of store and distance from customer address to the store. The parameter for distance shows the intention to disturb to access to store. Huff model was used in various scenes to analyze trade area or store power.

However, Huff model needs not only own store data but also competitive store around the store. Thus, when the manager of a chain wants to calculate his trade area, Huff model cannot utilize to this objective. Many of the other attraction models like MLN or MCI model are needed competitive stores' data. However, in real business situation, it is very difficult to obtain these other stores' data. Thus, it is an important issue to analyze the trade area or selling power of own store using only already getting data.

Yamazaki (1996) analyzed trade area using transaction data of a sport club. The result was displayed on geographic information system. The manager could grasp the trade area of the club by observing the display and plan an effective advertising like folding flyer of newspaper.

Yokoyama et al. (1996) showed trade area and purchase model to predict the amount of sales with consumers' preference based on Huff model. They used conjoint analysis to analyze consumer's preference, then regression model are utilized to predict the amount of sales. From the result of analysis, they pointed out that accessibility is one of the most attractive factor for choosing store.

In this study, we use the POS data with customer identifying data of the supermarket chain (i.e. ID-POS data) and some opened statistical data or GIS (geometrical information data) like Google map. The stores of the chain are located at the area of a Japanese regional urban area, and there are over 50 stores which are located in near prefecture. Some of them are located on plain but the others are located in mountainous area. The number of residence and competitive store around each store is not same, thus the competitive situation is not same, too. Then we need to consider these specific conditions to estimate the trade area.

## 3 Data

### 3.1 About Supermarket Chain

In this study, we focus on a Japanese supermarket chain. This chain have over 53 stores in a same region. We name each store from S1 to S53. Some of them are located on urban or suburbs, however the other is located on country area. The size of the largest store is about 2,000 m<sup>2</sup> however the smallest store is only 300 m<sup>2</sup>. These stores treat all categories of food mainly.

### 3.2 Data Summary

This chain introduce member card system (frequent shoppers program: FSP) and POS (point of sales) system. ID-POS (POS data with customer Identification number) contains purchase date, time, receipt number, purchase items, the number of purchase and price with customer ID. Thus, the manager can obtain detail purchase record of each customer. We use 3 months records (04/2015–06/2015), the summary statistics is shown in Table 1.

**Table 1.** Chain summary

| Term                      | From 01/04/2015 to 30/06/2015 |
|---------------------------|-------------------------------|
| No. of store              | 53 (located on 2 prefectures) |
| No. of member card holder | About 650 thousand            |
| No. of purchase items     | About 8.9 million items       |
| Purchase amount           | About 17.5 billion JPY        |

The store size and location cluster are shown in Tables 2 and 3, respectively.

**Table 2.** Store size categories

| Size                              | Store   |
|-----------------------------------|---|
| Small<br>( $<1,000 \text{ m}^2$ ) | S1, S2, S3, S4, S7, S9, S10, S11, S14, S15, S16, S20, S33, S34, S42, S45, S46, S53  |
| Medium<br>(other)                 | S5, S6, S8, S12, S17, S18, S19, S21, S22, S23, S25, S26, S27, S28, S29, S31, S32, S36, S37, S38, S43, S44, S47, S48, S49, S50, S51, S52 |
| Large<br>( $>1,500 \text{ m}^2$ ) | S13, S24, S30, S35, S39, S40, S41   |

**Table 3.** Location categories

| Area    | Store   |
|---------|---|
| Urban   | S1, S5, S9, S10, S19, S22, S25, S30, S35, S36, S39, S45, S51  |
| Suburbs | S2, S3, S4, S11, S12, S13, S14, S15, S16, S18, S20, S21, S23, S28, S31, S32, S33, S34, S37, S38, S40, S41, S42, S43, S44, S46, S47, S48, S49, S50, S52, S53 |
| Local   | S6, S7, S8, S17, S24, S26, S27, S29   |

Figure 1 shows the sales transition in analyzing term. As shown in Fig. 1, sales amount of almost weekend are higher than weekday, however especially Monday and Friday are typical lower. However, in early in May, it was continuing high amount due to long holiday (in Japan, it is calls Golden Week from late April to early May).

When we focus on each store, the amount of sales of each store must not have this common rule. Figure 2 shows the transition of sales amount of a store (S1). This store is small and located on urban area. We can guess that many of customers are office worker and residents near store, thus the sales is not high on weekend. Actually, this

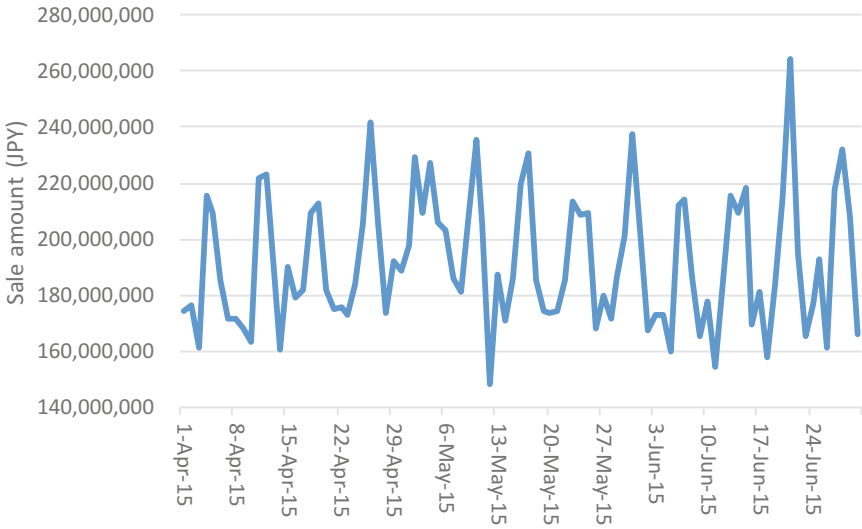


Fig. 1. Sales transition

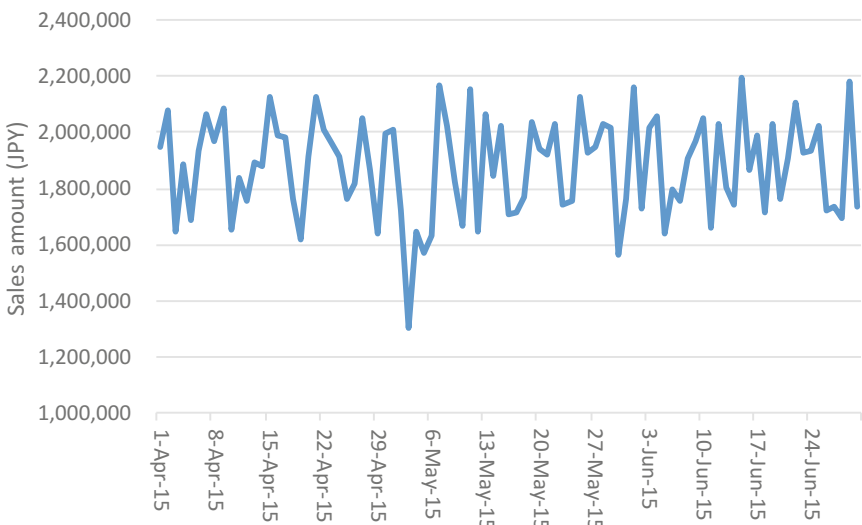


Fig. 2. Sales transition of S1

store emphasizes lunch box or side dishes on assortment rather than fresh items such as vegetables, meats or fishes. As shown in this example, the sales is not same among the stores.

Figure 3 shows scatter plots of the number of visit, number of purchase item and purchase amount for each customer and each store. The correlation between the

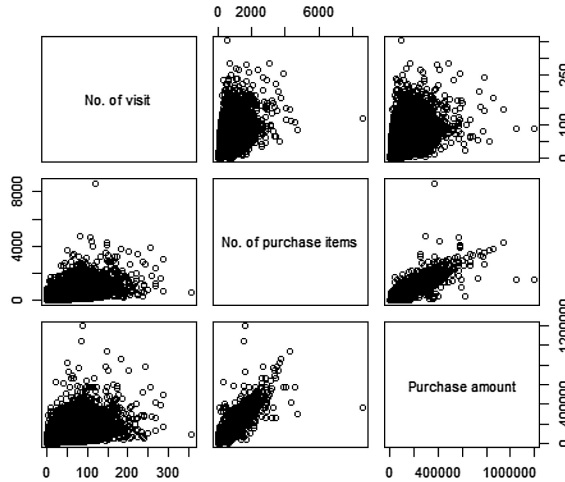


Fig. 3. Scatter plot of purchase data

number of purchase item and purchase amount is high, however the correlation between the number of visit to store and the other variables are not high. It may show that usage of store is not same with respect to each store.

### 4 Analysis and Discussions

The outline of our analysis is shown in Fig. 4. First we aggregate the purchase data with respect to each customer and each store, second calculate the radius of some percentile distance for each store, and third investigate the cause of effect to decide the radius.

The detail of our model is explained from the next subsection.

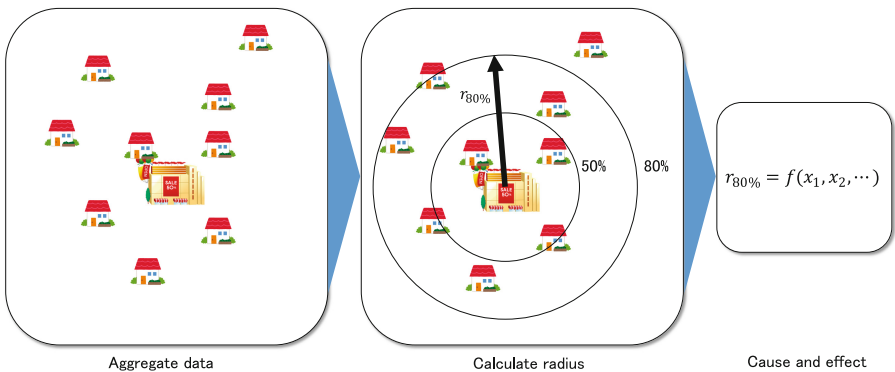


Fig. 4. Outline of our analysis

#### 4.1 Distance Between Customer and Store Address

To calculate the distance between each customer's address and store address, first, we give longitude and latitude to all addresses using google geocoding. Next, the distance is calculated using Hubeny's formula, which considers the curve of the earth in order to determine the distance between two coordination. The equation is given as follows,

$$d = \sqrt{(d_y R)^2 + (d_x N \cos \mu_y)^2} \quad (1)$$

where  $d_y$  is the latitude difference between two points,  $d_x$  is the longitude difference,  $\mu_y$  is the average latitude of the two positions,  $R$  is the radius of curvature of the meridian, and  $N$  is the transverse radius of curvature. The data were projected on the WGS84 datum. Thus, the following ellipsoid parameters were obtained: 6,378,137 m for the semi-major  $a$  and 6,356,752 m for the minor  $b$ . Moreover,  $R$  and  $N$  are defined as follows.

$$R = \frac{a(1 - e^2)}{W^3} \quad (2)$$

$$N = \frac{a}{W} \quad (3)$$

$$W = \sqrt{1 - e^2 \sin^2 \mu_y} \quad (4)$$

$$e = \sqrt{\frac{a^2 - b^2}{a^2}} \quad (5)$$

$R$  and  $e$  is called "meridian radius of curvature" and "major eccentricity", respectively.

#### 4.2 Calculate Radius of Trade Area

Next, we calculate the radius of trade area of each store.

To achieve this, first, we gather the purchase data of each store from database. Second, the number of visiting, the number of purchase item are summed up with respect to customer, then we put it in order by the distance from the store. Third, we find the distances of some cumulative probability. Table 4 shows an example. The 1<sup>st</sup> column is the distance from a store to customer's home address, and it is arranged by distance. From the 2<sup>nd</sup> to 4<sup>th</sup> columns are Number of visiting the store, the number of purchase items and purchase amount. Moreover, from the 5<sup>th</sup> to 7<sup>th</sup> of the table is cumulative ratios for each variable. As shown this table, we found 5 for 50 percentile distance (radius) according to the number of visiting. Thus we can interpret 50% of customer are into circle whose center are store with radius 5. Furthermore, we found 4 and 3 of distance for the number of purchase and purchase amount, respectively.

**Table 4.** An example to calculate radius

| Distance | No. of visit | No. of purchase | Purchase amount | Cum. no. of visit | Cum. no. of purchase | Cum. purchase amount |
|----------|--------------|-----------------|-----------------|-------------------|----------------------|----------------------|
| 1        | 4            | 65              | 1200            | 4.0%              | 6.5%                 | 10.0%                |
| 2        | 14           | 100             | 2000            | 18.0%             | 16.5%                | 26.7%                |
| 3        | 20           | 210             | 2800            | 38.0%             | 37.5%                | <b>50.0%</b>         |
| 4        | 4            | 125             | 1300            | 42.0%             | <b>50.0%</b>         | 60.8%                |
| 5        | 8            | 75              | 1200            | <b>50.0%</b>      | 57.5%                | 70.8%                |
| 6        | 9            | 125             | 900             | 59.0%             | 70.0%                | 78.3%                |
| 7        | 10           | 80              | 400             | 69.0%             | 78.0%                | 81.7%                |
| 8        | 4            | 70              | 500             | 73.0%             | 85.0%                | 85.8%                |
| 9        | 15           | 50              | 200             | 88.0%             | 90.0%                | 87.5%                |
| 10       | 12           | 100             | 1500            | 100.0%            | 100.0%               | 100.0%               |
| Total    | 100          | 1000            | 12000           |                   |                      |                      |

### 4.3 Result of Analysis

Using the method in the previous subsection, we can calculate the radius for each cumulative ratio. Table 5 shows the summary statistics of radius for some percentiles.

**Table 5.** Summary statistics of radius (meter)

| Percentile               |          | 25%    | 50%    | 70%     | 80%     | 90%     |
|--------------------------|----------|--------|--------|---------|---------|---------|
| No. of visiting          | min.     | 82.7   | 299.9  | 547.4   | 773.8   | 1161.1  |
|                          | average  | 522.3  | 1143.5 | 2196.3  | 3277.6  | 5577.7  |
|                          | median   | 381.9  | 766.5  | 1430.3  | 2282.2  | 4364.3  |
|                          | max.     | 1690.0 | 3952.8 | 10027.4 | 10938.2 | 17934.7 |
|                          | st. dev. | 383.5  | 877.1  | 2075.3  | 2761.3  | 3988.4  |
| Number of purchase items | min.     | 82.7   | 299.9  | 565.2   | 773.8   | 1116.6  |
|                          | average  | 571.6  | 1169.8 | 2201.6  | 3268.7  | 5271.2  |
|                          | median   | 409.4  | 788.6  | 1393.9  | 2282.2  | 4191.9  |
|                          | max.     | 1933.0 | 3952.8 | 10088.3 | 10938.2 | 17410.6 |
|                          | st. dev. | 429.5  | 909.3  | 2149.1  | 2854.7  | 3999.5  |
| Purchase amount          | min.     | 94.8   | 299.9  | 598.0   | 773.8   | 1116.6  |
|                          | average  | 582.6  | 1233.6 | 2244.8  | 3299.1  | 5346.9  |
|                          | median   | 473.6  | 788.6  | 1430.3  | 2340.7  | 4204.8  |
|                          | max.     | 1933.0 | 4834.8 | 10088.3 | 11647.6 | 18102.6 |
|                          | st. dev. | 423.8  | 1028.7 | 2214.1  | 2902.2  | 4081.7  |

As shown in Table 5, for example, the number of purchase items of 50% have various values, the minimum is about 300 m but the maximum is about 4,000 m. Figure 5 shows the histograms of 80% of the number of visiting. The radius of each store is not similar, because the location situation and customer are not same.

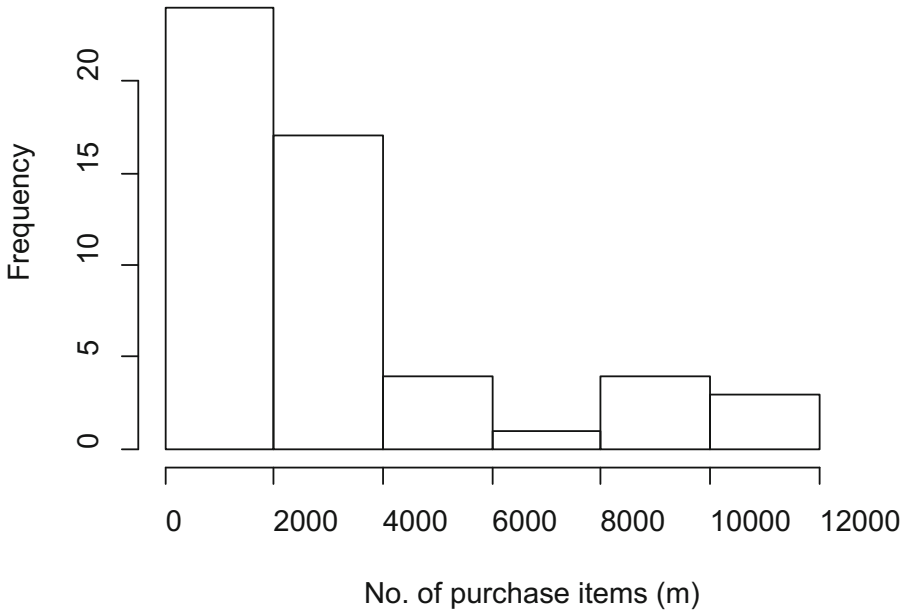


Fig. 5. Histogram of 80% trade area of the number of purchase

For the next analysis, we analyze the effect of various causes for trade area size. To do this, we use multiplicative regression analysis, the equation is shown in Eq. (6).

$$y_i = \beta_0 \prod_{j=1}^p \beta_j^{x_{ij}} \varepsilon_j \tag{6}$$

where  $\beta_0, \beta_1, \dots, \beta_p$  are the intercept and slope parameters and  $\varepsilon_j$  is the residual. Taking the logarithm of  $y_i$ , Eq. (6) can be treated as linear model which can be used the ordinary least square method. The reasons why multiplicative model is adapted is the distance is not negative value and effects of some variables seem exponentially, when we use multiplicative model, we obtain only positive predicted values. In addition, the variables vary broadly, thus if we use a linear model, the residual may not distribute homogeneously.

In this study, we use “No. of parking,” “square root of parking,” “cube root of parking,” “sales area,” “location,” “No. of items” “population around store (in 1 km radius of store address),” “No. of household (in 1 km radius of store address)” and “No. of household size” for explanatory variables of our regression model. The reason why we set 3 kinds of parking lot is the effectiveness of the number of parking does not seem linear, thus the combination of these variables may express non-linear effect. The variable “location” has 3 factors; urban, suburb or country, and the other variables are



continuous. In analysis, we omit “urban” level. The variables according with household are gotten from jSTAT MAP supported by National Statistics Center<sup>1</sup>.

The response variable is 80% cumulative ratio of “No. of visiting,” “No. of purchase items” and “purchase amount.”

When we analyze regression model, we adopt variable selection to choose significant variable statistically<sup>2</sup>. Table 6 shows the selected variables and the value of parameters for 3 kinds of 80% radius model. All results of our models are summarized in the tables of appendix.

**Table 6.** Result of multiplicative regression model (The value is displayed by exponential form, so for example E-01 means  $10^{-1}$ ).

| Explanatory variables  | No. of visiting | No. of purchase items | Purchase amount |
|------------------------|-----------------|-----------------------|-----------------|
| Intercept              | 8.39E+00        | 7.90E+00              | 7.81E+00        |
| No. of parking         | -1.16E-02       | -1.11E-02             | -1.08E-02       |
| Square root of parking | 9.78E-01        | 8.71E-01              | 8.39E-01        |
| Cube root of parking   | -2.02E+00       | -1.68E+00             | -1.60E+00       |
| Location (suburb)      | 3.31E-01        | 3.05E-01              | 3.04E-01        |
| Location3 (country)    | 1.21E+00        | 1.18E+00              | 1.18E+00        |
| Population             | -2.15E-05       | -2.07E-05             | -2.09E-05       |
| No. of household       | 3.67E-05        | 3.35E-05              | 3.38E-05        |

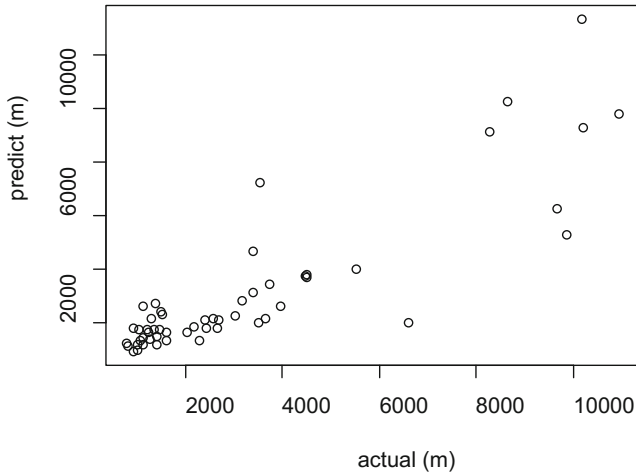
All 3 models of Table 6 selected common variables. When population is lower then the trade area is larger. We can interpret the result that a store located on country need customers who live far from the store, thus as a result, the trade area becomes wider.

The value of multiplicative correlation coefficients are about 0.870 for all models. Figure 6 shows the scatter plot of actual versus predicted value of trade area for the number of purchase items. Some points are not close to 45o line, thus we may need some further analysis. However, almost stores are well predicted, thus the results are generally appropriate. All scatter plots are shown in appendix in case of higher percentage, the multiple correlation coefficient value are higher, thus outliers are fewer and the prediction is achieved appropriately. However about especially larger distance, the other variables may be needed additionally.

About percentiles, models of higher percentile are better predicted. Especially, 25% cumulative probability model are not well-predicted. One of the reasons obtained these results, we may point out the core customer of store are not 25% region, but broader area. Moreover about the larger radius, the fitting of predicted value are not well. Almost of these store are located on country, thus the number of population around store is not many thus the stores must consider the wider area, thus the various noise (e.g. the number of competitive store or difference of lifestyle of residences) may be ignoring in our models. However, our model are well-predicted on the whole.

<sup>1</sup> <https://jstatmap.e-stat.go.jp/gis/nstac/>.

<sup>2</sup> We used “lm” and “step” function of R language in analyses. So the variable selection is based on AIC.



**Fig. 6.** Actual vs. predict (No. of purchase items)

## 5 Concluding Remarks

In this study, we focus on consumer behavior, especially selecting store behavior. From the viewpoint of store or store manager, we analyzed the range of consumer for the store that is trade area, using ID-POS data. Then we analyzed the trade area radius using regression model. From our result, we could estimate well prediction for trade area radius. Our model can be utilized to optimize area marketing strategy or open a new store, because the trade area can be estimated.

In this study, we only calculate the radius for each store, however, we did not consider items which is assortment in store. To match the needs of customer of each store, we need to know the true needs of core customer or heavy user. In future work, we will consider the needs of items and gather our analysis. In addition, we did not consider the competitive store around each store. When the number of competitor are larger, than the competition are severer, thus it may be effected on the change of trade area radius. Moreover, the coexistence of plural stores of same chain were not considered. Some customer may use properly some stores, for example, lunch box is purchased near office, however almost of foodstuff are purchased near home. These are also our future works.

## Appendix

In this appendix, we show the result of our multiplicative regression models. From Tables 7, 8 and 9 show the coefficient value of each data. Figure 7 shows scatter plots of actual versus predicted value obtained our regression analyses.

**Table 7.** Coefficient values of multiplicative regression model (No. of visiting)

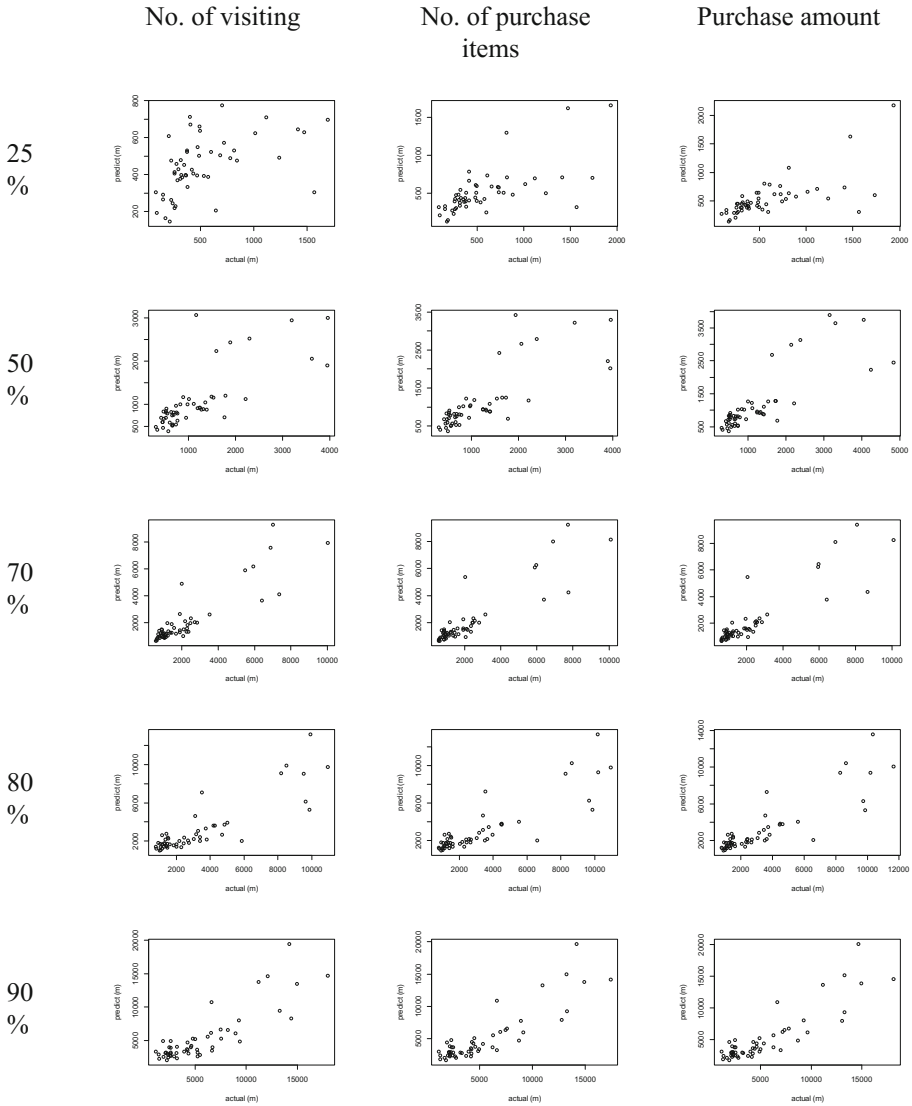
| Variables                        | 25%       | 50%       | 70%       | 80%       | 90%       |
|----------------------------------|-----------|-----------|-----------|-----------|-----------|
| Intercept                        | 4.44E+00  | 5.89E+00  | 7.76E+00  | 8.39E+00  | 9.35E+00  |
| No. of parking                   | -3.77E-03 | -1.93E-03 | -8.99E-03 | -1.16E-02 | -1.15E-02 |
| Sq root of No. of parking        | 1.47E-01  | 9.23E-02  | 8.18E-01  | 9.78E-01  | 1.02E+00  |
| Cube root of No. of parking      |           |           | -1.74E+00 | -2.02E+00 | 2.19E+00  |
| Sales area                       |           |           |           |           |           |
| Location (suburb)                |           | 3.17E-01  | 4.21E-01  | 3.31E-01  | 3.21E-01  |
| Location (country)               |           | 1.07E+00  | 1.45E+00  | 1.21E+00  | 1.08E+00  |
| No. of items                     | 3.10E-05  | -2.39E-06 |           |           | -2.23E-05 |
| Population                       |           |           | -1.54E-05 | -2.15E-05 | 4.04E-05  |
| No. of household                 |           |           | 2.81E-05  | 3.67E-05  |           |
| No. of person per household      | -7.52E-06 |           |           |           |           |
| Multiple correlation coefficient | 0.499     | 0.764     | 0.890     | 0.867     | 0.855     |

**Table 8.** Coefficient values of multiplicative regression model (No. of purchase items)

| Variables                        | 25%       | 50%       | 70%       | 80%       | 90%       |
|----------------------------------|-----------|-----------|-----------|-----------|-----------|
| Intercept                        | 3.47E+00  | 5.81E+00  | 7.19E+00  | 7.90E+00  | 9.12E+00  |
| No. of parking                   |           | -2.33E-03 | -7.04E-03 | -1.11E-02 | -1.24E-02 |
| Sq root of No. of parking        | 4.77E-02  | 1.09E-01  | 5.87E-01  | 8.71E-01  | 1.06E+00  |
| Cube root of No. of parking      |           |           | 1.13E+00  | -1.68E+00 | -2.23E+00 |
| Sales area                       |           |           |           |           |           |
| Location (suburb)                |           | 3.07E-01  | 2.09E-01  | 3.05E-01  | 3.37E-01  |
| Location (country)               |           | 1.10E+00  | 1.32E+00  | 1.18E+00  | 1.14E+00  |
| No. of items                     | 5.13E-05  |           |           |           | -2.03E-05 |
| Population                       | -2.56E-06 | -2.91E-06 | -2.11E-06 | -2.07E-05 | 3.51E-05  |
| No. of household                 |           |           |           | 3.35E-05  |           |
| No. of person per household      | 4.40E-01  |           |           |           |           |
| Multiple correlation coefficient | 0.670     | 0.805     | 0.899     | 0.870     | 0.878     |

**Table 9.** Coefficient values of multiplicative regression model (Purchase amount)

| Variables                        | 25%       | 50%       | 70%       | 80%       | 90%       |
|----------------------------------|-----------|-----------|-----------|-----------|-----------|
| Intercept                        | 3.99E+00  | 5.81E+00  | 7.17E+00  | 9.12E+00  | 9.11E+00  |
| No. of parking                   |           | -2.33E-03 | -7.03E-03 | -1.24E-02 | -1.24E-02 |
| Sq root of No. of parking        | 3.99E-02  | 1.09E-01  | 5.77E-01  | 1.06E+00  | 1.06E+00  |
| Cube root of No. of parking      |           |           | -1.09E+00 | -2.23E+00 | -2.22E+00 |
| Sales area                       | 6.23E-04  |           |           |           |           |
| Location (suburb)                |           | 3.07E-01  | 1.96E-01  | 3.37E-01  | 3.39E-01  |
| Location (country)               |           | 1.10E+00  | 1.30E+00  | 1.14E+00  | 1.12E+00  |
| No. of items                     |           |           |           |           |           |
| Population                       | -3.40E-06 | -2.91E-06 | -2.43E-06 | -2.03E-05 | -2.11E-05 |
| No. of household                 |           |           |           | 3.51E-05  | 3.65E-05  |
| No. of person per household      | 5.31E-01  |           |           |           |           |
| Multiple correlation coefficient | 0.687     | 0.825     | 0.892     | 0.878     | 0.878     |



**Fig. 7.** Scatter plot of actual vs. predicted radius (horizontal axis = actual, vertical axis = predicted)

**References**

Huff, D.L.: A probabilistic analysis of shopping center trade areas. *Land Econ.* **39**, 81–90 (1963)  
 Hubeny, K.: Weiterentwicklung der Gauss’schen Mittelbreitenformeln. *Z. Vermess* **84**, 159–163 (1959)  
 Lilien, G.L., Rangaswamy, A., De Bruyn, A.: *Principles of Marketing Engineering*. 2nd edn. DecisionPro (2012)

- Sato, E.: Present and tasks of trade area analysis. *Commun. Oper. Res. Soc. Jpn.* **42**(3), 137–142 (1997). (in Japanese)
- Yamazaki, T.: An application of geographic information system to the trade area analysis for commercialized sports clubs. *J. Jpn. Soc. Sports Indus.* **6**(2), 15–23 (1996). (in Japanese)
- Yokoyama, S., Okanda, H., Ishikawa, K.: Market area analysis in new general merchandise store advancement project. *J. Soc. Proj.* **2**(2), 19–26 (1996)