

# THE MULTIPLE-LINEAR REGRESSION MODEL FOR ANALYZING THE FACTORS OF CUSTOMER'S PURCHASING PROCESS

Eri Usami<sup>1</sup>, Vajirasak Vanijja<sup>2</sup> and Masahiro Inoue<sup>1</sup>

<sup>1</sup>Graduate School of Engineering and Science, Shibaura Institute of Technology

<sup>2</sup>School of Information Technology, King Mongkut's University of Technology Thonburi

Contact E-mail Address: mf13014@shibaura-it.ac.jp

**ABSTRACT** In recent years, big data is attracting attention. In retail stores, Point of Sales (POS) system became popular to record and analyze sales information. We focused on two type data, not only after-sales data but also before-sales data (products touched history before purchasing). In this study, we collected and analyzed the sales data and touched data for a clothes shop. The purpose of this study to analyze the factors which affect the number of sales and the number of touches, moreover we discuss the relationship between sales data and touched data. We conducted gender-segregated two experiments in a simulated environment for Thai university's 34 students. We collected sales and number of touched data by this experiment. For the second experiment, we set the product property scores by questionnaire. Then we analyzed customer's behavior data and product property score data by using the multiple linear regression analysis. As a result of analyzing, there is a correlation between sales and touched number. The most affected factor for boys is the design of product and for girls the position displayed in the store.

## 1. INTRODUCTION

### 1.1 Back ground

In recent years, big data is attracting attention. In retail stores, Point of Sales (POS) system became popular to record and analyze sales information. However, the POS system records only after-sales data. We don't know the purchasing process of customers. We focused on two type data, not only after-sales data but also before-sales data (products touched data before purchasing). Previous study focused on products purchasing process behavior of customers and collected data using RFID or camera (Takata & Inoue, 2011, Araki, 2009). However those

studies didn't have enough analysis. The other study aimed to clarify the products features that are related with the time of hesitation during customer's products contact process. The product contact data, which show the customer's products contact process, shelf rate, brand, function, and price in shops, are observed, and multiple linear regression analysis was conducted (Nishina, 2012).

### 1.2 Purpose

In this study, we collected and analyzed the sales data and touched data for a clothes shop. The purpose of this study to analyze the factors which affect the number of sales and the number of touches, moreover we discuss the relationship between sales data and touch data.

## 2. EXPERIMENT

### 2.1 Experiment

We conducted gender-segregated two experiments in simulated environment for Thai university's 34 students. Those experiments were carried out to collect customer's behaviors data and to set the product property scores. First, students shopped on the website for experiment. And sales and number of touched data collected in this experiment. Products thumbnails and price are listed on web page. If shoppers are interested some products, the shoppers can see the products details when they click on the thumbnail image. In this study, this number of click means number of touched. For the second experiment, we set the product property scores by questionnaire: "Is this price (design) good for you?". The product properties scores consist of price, design and position. This questionnaire is five-grade evaluation. As for position score, products that put at the top of the line get 5 points and the lowest line product get 1 point.



Fig. 1 Website for experiment

## 2.2 Analysis method

We use the multiple-linear regression model for analysis. It is popular to analyze what factors affect sales and to predict the sales from objective variable as the weather, the time, and floor space. In this study, the dependent variables are sales and number of touched. The independent variables are product property scores (price, design and position).

$$y = ax_1 + bx_2 + cx_3 + d \quad (1)$$

## 3. RESULT

### 3.1 Analysis Results

Figure 2 and 3 show the relationship between sales data and touched data. The vertical axis indicates the number of sales and the horizontal axis indicate the number of touched.

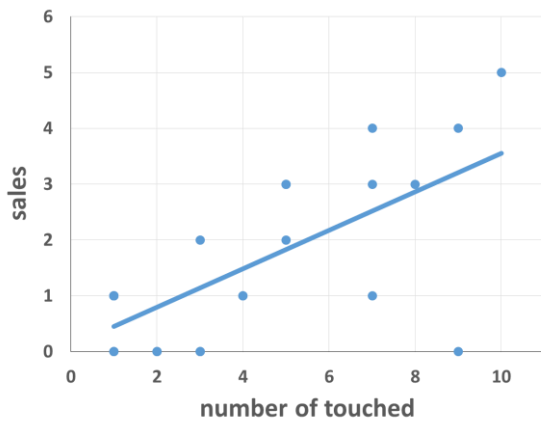


Fig.2 Relationship between sales data and touched data (Boys)

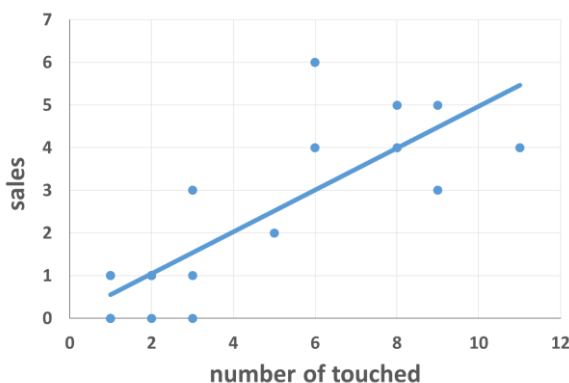


Fig.3 Relationship between sales data and touched data (Girls)

Table 1 and 2 show the result of multiple regression analysis.

Table 1 (Boys) sales & touched

	Coefficient	t	P-		Coefficient	t	P-
Price	0.468086	0.57632	0.570829	price	0.135651	0.069219	0.945503
design	2.06946	2.262782	0.034933	design	3.05521	1.384494	0.181461
position	-0.08997	-0.49488	0.626079	position	0.29392	0.670062	0.510481

Table2 (Girls) sales & touched

	Coefficient	t	P-		Coefficient	t	P-
Price	-1.2315	-1.73531	0.100777	price	-2.05966	-1.53211	0.143892
design	1.148328	1.683072	0.110634	design	2.532869	1.959757	0.066629
position	0.712263	2.914536	0.00966	position	0.740006	1.598517	0.128347

For the left half in both tables, the dependent variable is sales, and for the right half, the dependent variable is number of touched. In the tables, “t” mean t-value: the higher values, the bigger effect. In general, effective variable is when t-value is bigger than 2. “P-” mean P-value: the lower values, the more trust.

### 3.2 Discussion

The result of analysis indicates a positive correlation between sales and touched number. Most products with large touched number had large number of sales. And most them with small touched number had and small number of sales. For both sales and number of touched, the most affected factor for boys was the design. For sales, the most affected factor for girls was the position. Then design is the most affected for number of touched, however there is not much difference between design and position for number of touched result. Altogether, it seems that position score is the most affected for girls.

## 4. CONCLUSION

This study focuses on the before-sales data (products touched data before purchasing). We collected and analyzed the sales data and touched data for a clothes shop. We conducted the experiment in a simulated environment for Thai university’s 34 students. We analyzed customer’s behavior data using the multiple linear regression analysis. Independent variables are three variables: price, design and position in this analysis. As a result of analyzing, there is a positive correlation between sales and touched number. The most affected factor is the design for boys and the position displayed in the shop for girls.

## REFERENCES

- Araki, T., Yonezawa, T., Nakazawa, J., Takashio, K. and Tokuda, H., IEICE Engineering Sciences Society, The Institute of Electronics, Information and Communication Engineers, *IEICE Technical Report*, 109(338), pp.13-18, 2009.
- Nishina, R., Suwa, H., Ogawa, Y. and Ohta, Y., The

Japan Society for Management Information, *Abstracts of Annual Conference of Japan Society for Management Information 2012f*, pp.123-126, 2012

Takata, K. and Inoue, M., The 15<sup>th</sup> IEEE International Symposium on Consumer Electronics, *Consumer Electronics (ISCE) 2011 IEEE 15th International Symposium on*, pp.458-461, 2011.



**Eri Usami** received B.E. degree from Shibaura Institute of Technology in 2013. She is currently a graduate student of Graduate School of Shibaura Institute of Technology.



**Vajirasak Vajirasak Vanijja** received his BE in Computer Engineering and MS in Information Technology from King Mongkut's University of Technology, Thonburi, Thailand(KMUTT) and received his Ph.D. in Computer Science from Japan Advance Institute of Science and Technology (JAIST). Currently he is an assistant professor at the School of Information Technology at KMUTT, and currently is director of the IP Communication Laboratory.



**Masahiro Inoue** received his B. S. and M. S. degree in Applied Physics from Waseda University and Ph. D. degree in Computer Science from Shizuoka University. He was engaged in research and development at Mitsubishi Electric Corporation. Since 2005 he has been a professor of Shibaura Institute of Technology. He is a Professional Engineer in Japan and a Project Management Professional. Current research interest: Embedded Systems Engineering and Project Management