<u>Project Report</u>

Conjoint Analysis
with
Dummy variable regression model

ETM 565/665

Research Method

for Engineering and Technology Management

Spring 2011

Dong-Joon, Lim

Dinesh, Soysa

Table of Contents

I. INTRODUCTION

There exist two types of approach statistical models are taking; 'compositional' and 'decompositional.' Models belong to former approach tries to explain the relationships among the existing data so that predict one's value from changes of others whereas latter tries to explain the relationships by assigning from one's given value to others. To illustrate this difference, we chose two typical models for both approach; regression model and conjoint analysis, and compare them focusing on their inner process. Especially, to explain the conjoint analysis, illustrative model for designing ideal smart phone is provided with step by step procedures in SPSS. Finally, idea incorporating advantages of both compositional and decompositional model is proposed for the future research.

II. CONJOINT ANALYSIS

A. What is it?

Conjoint analysis, a statistical technique created by mathematical psychologists from 1970 [1], is used to describe a broad range of techniques to estimate how much people value certain attributes and features of a new product or service. Pair wise comparison, Tradeoff Matrices, Discrete Choice, and Hierarchical Choice are few examples of various types of conjoint analysis [2]. It is widely used specially by the marketers to understand the way buyers will differentiate the values of different attributes and features of a product or service. This method can be a very powerful tool for businesses when it comes to understanding the customer needs and designing user centered products or services that meet customer needs.

The objective of conjoint analysis is to determine which attributes is most influential on respondent choice or decision making. Based on this data, markets can focus on the most important features of products or services and design them accordingly to gain the attention of its target buyers [3].

A major advantage of using conjoint analysis over the traditional survey methods is that conjoint analysis challenges to break the task of estimate into a series of choices or ratings. Traditional methods on the other hand, ask the respondent to estimate the value they will assign for each attribute. Another advantage of using conjoint analysis is that result can be used to develop market simulation models that can be very useful in the future. This is especially true with the ever changing markets.

B. How it works?

After conducting a conjoint analysis depending on its type, statistical methods like weighted least squares regression, ordinary least squares regression and logit analysis will be used to translate the answers of the respondent into important values or utilities. The real values obtained by these statistical methods however are not important. Only the relationships or relative values between each of the attributes are necessary. The main purpose of these calculations is to understand the respondent's answers in a way that describes the underlying value that they knowingly or unknowingly place on each attribute.

1) Steps in Conjoint Analysis

First marketer has to determine which product attributes or features are most important to the market [4]. Then they have to determine which data collection method to use in order to collect respondents data. They also need to figure out how these data will be captured simultaneously. The third step is to determine one of the today's most commonly used Choice-based conjoint or preference-based conjoint. Here, they need to select which conjoint method will best fit the research issue. The next step is to start collecting the data from the target market after pre-testing the attribute list and survey instruments. Then they have to calculate the utilities for each and every respondent. Finally, they have to create the market simulation model which later helps to predict the impact on introduction of new products and changes in existing products on the market [5].

2) How to define the attributes

When it comes to selecting attributes, it is very important to select them cautiously. Too many attributes and too few attributes can greatly reduce the accuracy of the conjoint model. If there are too many attributes, it will increase the burden on respondent. Simultaneously, too low attributes will cause missing information on the model. Therefore, it is recommended to conduct a qualitative research to create the list of key attributes for the product. After carefully selecting the product attributes, it is extremely essential to select the attribute levels.

For an example, product like a car can have attributes like maximum speed, gas mileage, and number of doors. These individual attributes have it is own attribute levels. For an attribute like maximum speed, the attribute levels would be specific speed points like 120mph, 140mph, and 160mph. Therefore it is very important to specify all the attributes and attribute levels of the product to create an accurate simulation model.

3) Overall value of the product

To calculate the overall product value, first we have to have utilities for each and every attribute level. Once we have this calculated, then we can add the utilities across all the attributes to calculate the product value for that product. Likewise, we have to repeat this process for every attribute in the conjoint model. Then finally we add the utilities for each attribute to calculate the total utility for that product.

C. Illustrative model

To identify the most important attribute of a new smart phone design, we have done a research and gathered all the necessary data of the last year's top 10 best-selling smart phone models. After carefully observing those data, we have identified display size, operating system, and carrier as the most important attributes. Then, we were able to identify three attribute levels for each and every attribute that we have previously identified. Those were;

Display Size = {3.0~3.5 inch, 3.5~4.0 inch, 4.0~4.5 inch}
Operating System = {iOS, Android, BlackBerry}
Carrier = {AT&T, Verizon, Sprint}

Next, we have created a table with all the data and then developed a program in SPSS to calculate the utility values for each level. The Fig. 1 shows the dataset that we have used.

	DS	OS	CR	STATUS_	CARD_
1	3.00	3.00	1.00	0	1
2	1.00	2.00	3.00	0	2
3	3.00	1.00	3.00	0	3
4	1.00	3.00	2.00	0	4
5	2.00	3.00	3.00	0	5
6	3.00	2.00	2.00	0	6
7	2.00	2.00	1.00	0	7
8	2.00	1.00	2.00	0	8
9	1.00	1.00	1.00	0	9

Figure 1 Dataset for Conjoint analysis

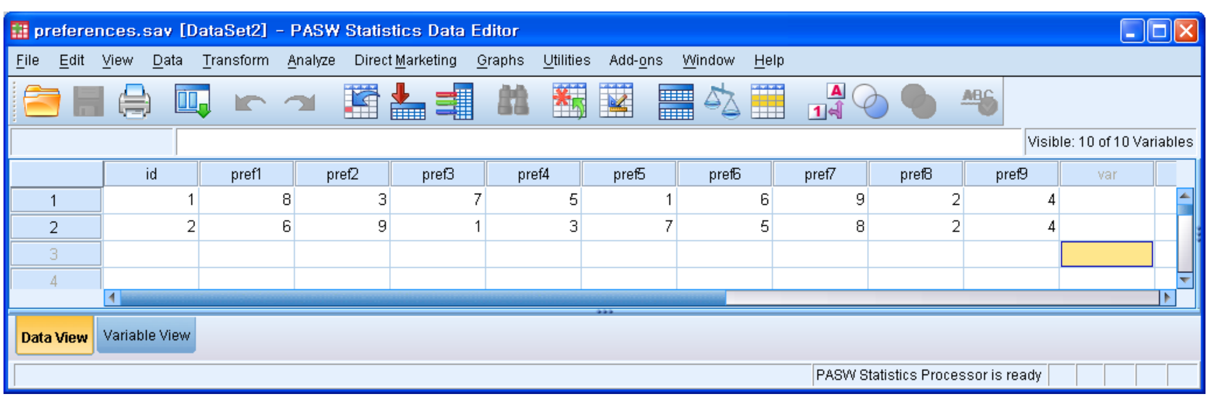
Since we have three attributes and each attribute has three attribute levels mathematically there are 27 (3*3*3) possible combinations. However, from Orthogonal design (Latin square method) only 9 (27/3) combinations will be prioritized. After run the dataset on SPSS the Card List was as shown below;

Card List

	Card ID	Display Size	Operating System	Carrier
1	1	4.0~4.5 inch	Blackberry	AT&T
2	2	3.0~3.5 inch	Android	Sprint
3	3	4.0~4.5 inch	iOS	Sprint
4	4	3.0~3.5 inch	Blackberry	Verizon
5	5	3.5~4.0 inch	Blackberry	Sprint
6	6	4.0~4.5 inch	Android	Verizon
7	7	3.5~4.0 inch	Android	AT&T
8	8	3.5~4.0 inch	iOS	Verizon
9	9	3.0~3.5 inch	iOS	AT&T

Figure 2 Sample profile from orthogonal design

After we have the Card List our next step was to rank the relative importance of each combination and create the preference dataset. The Fig. 3 shows the preference dataset.



	id	pref1	pref2	pref3	pref4	pref5	pref6	pref7	pref8	pref9	var
1	1	8	3	7	5	1	6	9	2	4	
2	2	6	9	1	3	7	5	8	2	4	
3											
4											

Figure 3 Preference evaluation

Once we have the reference dataset we have created a program to perform conjoint analysis on our dataset. Fig. 4 shows the program that we have created.

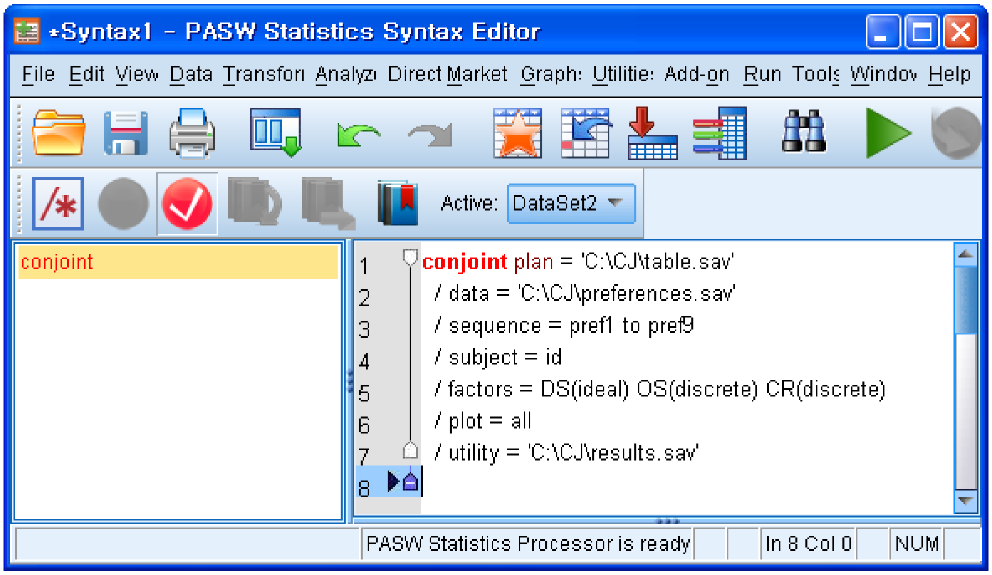


Figure 4 Codes for running conjoint analysis

According to the results of conjoint analysis we have found that Display Size is the most important attribute of new smart phone. Fig. 5 illustrates the average importance over the factor.

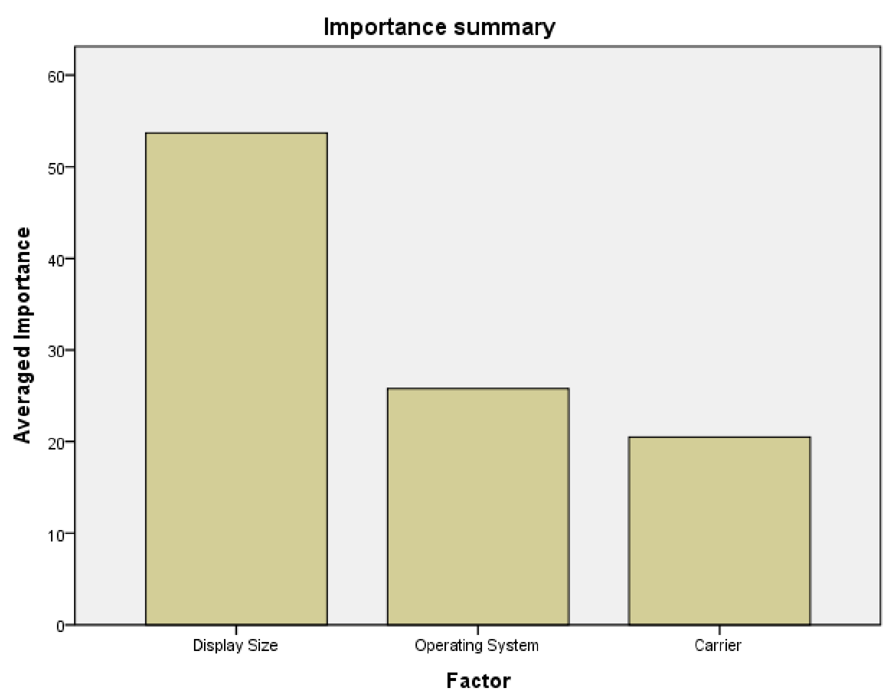


Figure 5 Importance of each attribute

III. Compositional model versus Decompositional model

A. Comparison with regression model

In compositional model, the researcher collects subject judgment from respondents on each element and then relates these ratings to overall preference in order to develop a predictive model [6]. One of the most typical compositional models is the regression analysis. As seen in the Fig. 6, regression analysis tries to explain the relationship between independent variable(s) and dependent variable(s). In other words, it calculates variance(s) of dependent variable(s) in terms of variance(s) of each independent variable(s). For example, in the Fig. 6, α indicates the averaged change of Y when X_1 increases or decreases assuming that other independent variables are held fixed. This is the reason why X_i is called independent variable(s) and multicollinearity is one of the biggest concerns in multiple regression analysis.

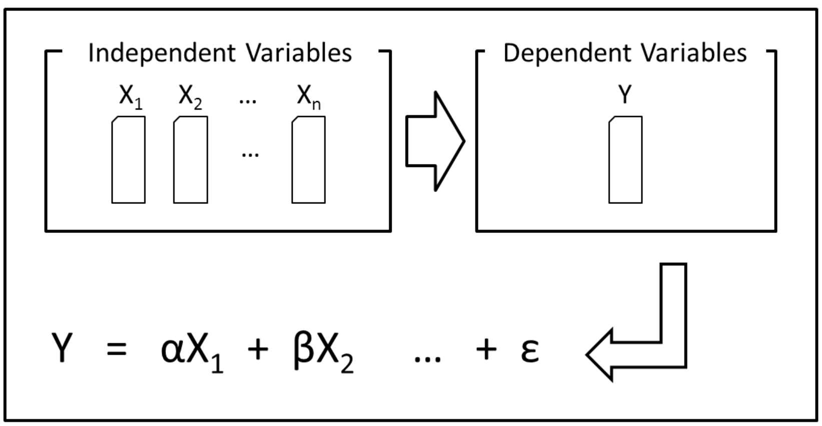


Figure 6 Regression model (Compositional model)

On the other hand, decompositional model just requires respondents' overall preference on sample profiles. It has been designed to be able to calculate relative degrees of each element with those input data. Literally, it decomposes the preferences of sample profiles and assigns it to each component (see Fig. 7). Hence, it can explain how much one element has impact on the overall preference relative to the others. For example, the height of L_{11} in Fig. 7 indicates the averaged preference of L_{11} when it presented in the sample profile. Not only that, by comparison of the difference between highest utility and lowest utility, we can interpret that which attribute is more important or how much overall preference is sensitive to each attribute. However, unlike regression model, it doesn't provide any information about variation of overall preference when level is added or replaced one to another.

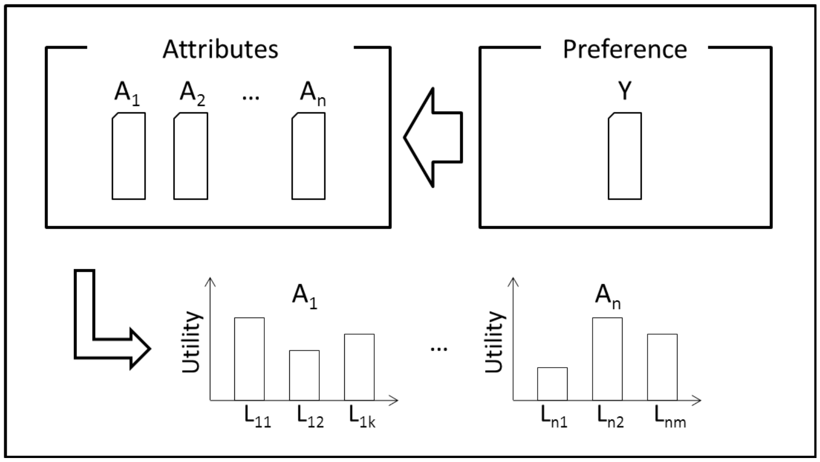


Figure 7 Conjoint analysis (Decompositional model)

Put geometrically, regression model is presented as following Fig. 8. Assuming that there is a distribution of dots in three dimension space formed by two independent variables (X_1 and X_2) and one dependent variable (Y), green line would be drawn as a regression curve. Here, we can interpret the regression equation as projecting green curve to the plane of each independent variable with dependent variable. That is, the red line is presenting the relationship between independent variable X_1 and dependent variable Y. Likewise, the blue line is presenting the relationship between independent variable X_2 and dependent variable Y. Decision maker may be able to figure out the optimum level of each variable or the efficient way to calibrate each variable for necessary action.

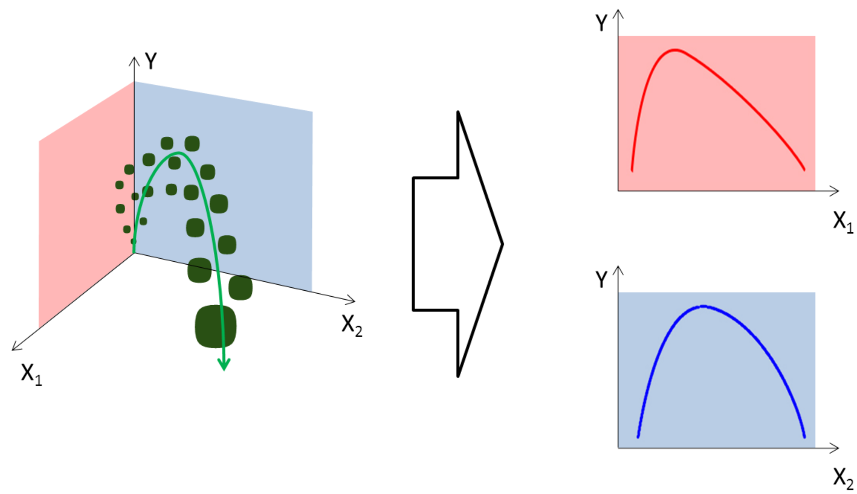


Figure 8 Regression model (geometric conceptualization)

Conjoint analysis, on the other hand, can be presented as following Fig. 9. Unlike regression model, attributes consist of discrete levels which are pre-defined in accordance with decision objective. And then, the number of profile is determined as grid in the figure. Each intersection indicates feasible sample profile which is to be prioritized by evaluators. Therefore, the green dots are plotted with the height of their preferences. Conjoint analysis, then, calculates the averaged utility of each level-coordinate. In this example, each level has presented in three sample profiles and been assigned their heights (preferences). Hence, red and blue bar charts indicate the average preference (utility) of each level when they presented in the sample profile. Note that these part-worth are relative utilities scaled to sum to zero within each attribute [7]. That is, level which shows negative utility value may have been very acceptable to all respondents. The metric in utility distribution is different from the rating used in the profile assessment that it doesn't represent the attractiveness [8]. In addition, these part worth are scaled as interval data. Hence, we cannot directly compare values between attributes even if they are showing same degree of utilities. And even though we are comparing utilities within the same attribute, we cannot say that one level is certain times as preferred as other because interval data do not support ratio operations. In the same manner, it doesn't explain how much the overall preference would increase or decrease when certain level is replaced by others.

Most common mistake when researchers interpret the result of the conjoint analysis is that they try to compose these utilities. They believe that the best combination will be made by combining the highest levels. For example, one could think that L_{12} and L_{22} would be the most effective combination based on the part worth results in Fig. 9. However, this notion is proved out to be false from the figure indicating that combination of L_{13} and L_{22} shows the highest preference among the sample profiles. The reason of utility L_{13} is lower than L_{12} is that it is an 'averaged value.' As shown in the figure, three green dots having L_{13} shows relative high deviation thereby making the highest preference offset by others. While it hides between other low values, part worth of L_{12} becomes the highest with small variance. This pitfall often brings about wrong decision especially when partial sample profile was assessed by orthogonal design. It should be stressed that the conjoint analysis only decompose the preference on sample profiles and present them as part worth distributions thereby making it possible to recognize the importance of each attribute. It doesn't guarantee the compositional process.

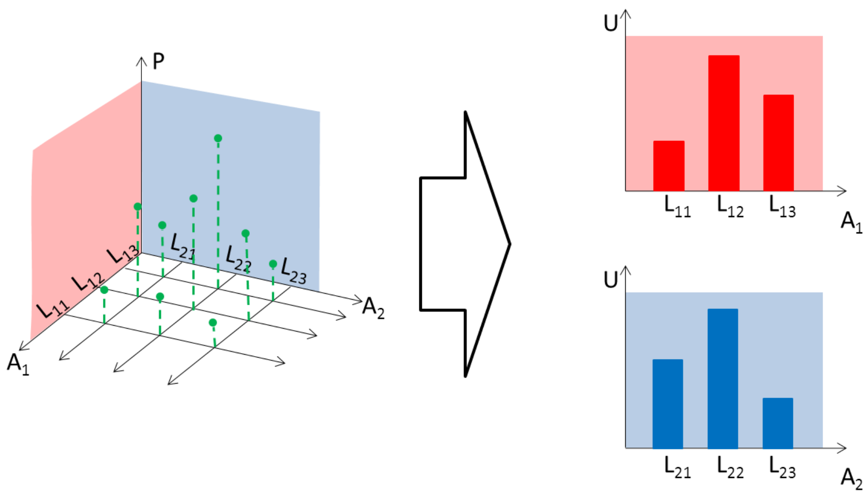


Figure 9 Conjoint analysis (geometric conceptualization)

B. Dummy variable regression model

Difference between continuous calibration of normal regression model and discrete levels of conjoint analysis can be made up by dummy variable regression model. In other words, dummy variable regression model can emulate part worth computation of conjoint analysis by defining its independent variables as 0-1 dummy variables. Fig. 10 illustrates this process with sample example which has two attributes and 6 levels. Because each attribute has three levels, possible 9 profiles can be presented with four dummy variables. For instance, a sample profile of L_{11} and L_{21} is coded as 1-0-1-0 respectively for each X_i variable. In the same way, a sample profile of L_{13} and L_{23} is coded as 0-0-0-0 for dummy variable X_i. As conjoint analysis, the value of Y indicates the preference of each sample profile. Finally, these data are to be run in dummy variable regression model with X_i as independent variables and Y as a dependent variable.

After running the regression model, we can interpret coefficients of the regression equation as utilities of each level [9]. In this example, α indicates the averaged utility of L_{11} when it presented in sample profiles. Likewise, β indicates the averaged utility that L_{12} has gotten from sample profiles. To make these values as pure conjoint utilities ($\hat{\alpha}$ and $\hat{\beta}$), it needs to be modified in order to satisfy the condition of sum to zero. We now have three equations; difference of utility L_{11} and L_{13} is equal to α, difference of utility L_{12} and L_{13} is equal to β, and sum of utility L_{11}, L_{12}, and L_{13} is equal to zero. By solving this simultaneous equation, in the end, utilities of three levels, $\hat{\alpha}$, $\hat{\beta}$, and -($\hat{\alpha} + \hat{\beta}$), are obtained. In the same way, utilities of levels belong to attribute 2 (A_2) can be calculated as $\hat{\gamma}$, $\hat{\delta}$, and -($\hat{\gamma} + \hat{\delta}$).

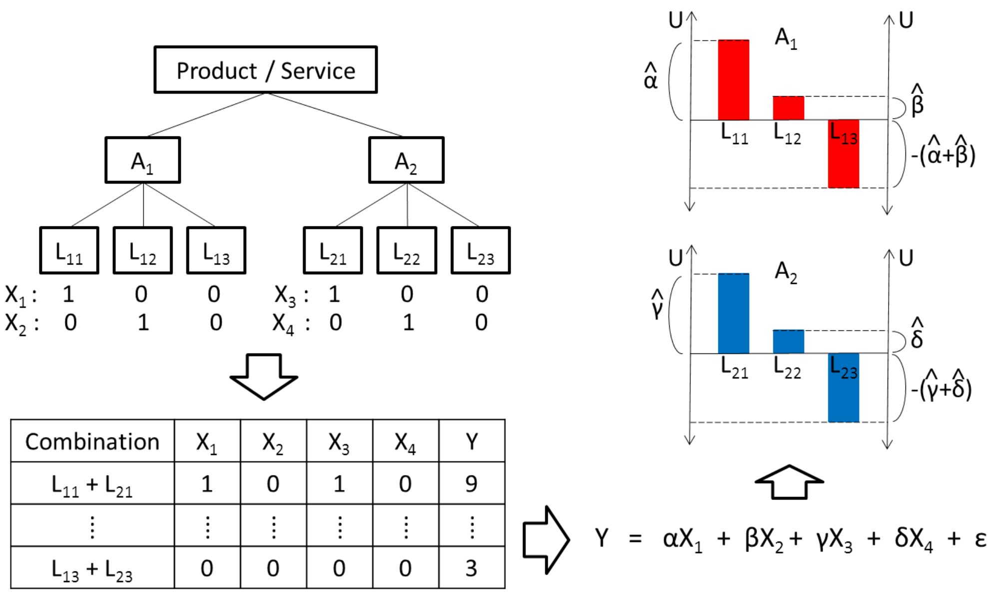


Figure 10 Dummy variable regression model

IV. CONCLUSION AND FUTURE RESEARCH

Conjoint analysis is a useful method in that it can elicit subject judgments on sub elements only with the overall preferences. This is the reason why the conjoint analysis is classified as one of decompositional models and widely used for product/service design. Regression model, on the other hand, is a typical example of compositional model. It requires data or evaluation on each variable so as to predict the change of certain variables as well as explain the relationship among variables. Interestingly, these two methods have complementary weaknesses each other due to their opposite characteristics. Conjoint analysis has a limitation that it doesn't guarantee the best combination of levels. In regression model parlance, it is not enough to predict the ideal point of the dependent variable only with part worth of each level. On the other hand, without either independent variables or dependent variables, regression model can neither estimate the other variable nor identify the relationships among the variables.

Recently, various network analysis models are actively developed and advanced in order to cope with interdependence problem that traditional model couldn't or just ignored. However, very worthwhile critiques on these methods commonly indicate that they put excessive evaluation burden on experts. Most of them basically require evaluators to fill out n^2 matrix for each decision criterion where n indicates the number of elements to be considered. No matter how these methods have been firmly designed to deal with interdependencies, this tedious procedure to gain input data which must be preceded sometimes makes it impractical to be implemented on real problem solving. Hence, setting aside creative algorithms to process data, the issue of gathering data itself needs to be taken into account as another research topic. Here, decompositional model seems to have potentials enough to cope with this issue. Its nature of making people evaluate profile rather than elements and then assigning preferences to them could help bridge this gap. In addition, existing decompositional procedure needs to be strengthened by incorporating various multivariate analyses such as combinatorial approach or MDS (Multi-Dimensional Scaling) to ensure validity of compositional functions toward finding the best combination with its part-utilities.

REFERENCE

[1] M. Halme and M.Kallio, "Estimation methods for choice-based conjoint analysis of consumer preferences," *European Journal of Operational Research*, vol 214, pp 160-167, 2011.

[2] R. Sethuraman, R. A. Kerin and W. L. Cron, "A field study comparing online and offline data collection methods for identifying product attribute preferences using conjoint analysis," *Journal of Business Research*, vol 58, pp 602-610, 2003.

[3] W. L. Moore, "Cross-validity comparison of rating-based and choice-based conjoint analysis models," *International Journal of Research in Marketing*, vol 21, pp 299-312, 2004.

[4] T. Ruetzler, J. Taylor, D. Reynoils, and W. Baker, "Understanding perceptions of professional attributes using conjoint analysis," *International Journal of Hospitality Management*, vol 30, pp 551-557, 2011.

[5] Webinar. (2011, June 3). *Utilizing Conjoint Analysis* [Online]. Available: https://www.agecon.purdue.edu/cab/programs/Webinar/Conjoint_Analysis/

[6] A. Sahni et al. (2011, June 3). *Conjoint Analysis* [Online]. Available: http://www.slideshare.net/madhusudanpartani/conjoint-analysis-3566217

[7] K. Ogawa, "An approach to simultaneous estimation and segmentation in conjoint analysis," *Marketing Science*, vol. 6, 1987.

[8] B. Orme, "Interpreting Conjoint Analysis Data," Sawtooth Software, Inc. Research Paper Series, 2002.

[9] Solvermark. (2011, June 3). *How to use dummy variable regression in excel to do conjoint analysis* [Online]. Available: http://www.squidoo.com/how-to-use-dummy-variable-regression-in-excel-to-do-conjoint-analysis