

A New Test of Attribute Significance for Nonparametric Conjoint Models

컨조인트 모형의 속성 유의성을 검증하기 위한 새로운 비모수통계 검증법

Hahn, Minhi(한 민 희)*
Krishnamurthi, Lakshman**
Kang, Hyunmo(강 현 모)***
Hyun, Jin-Seok(현 진 석)****
Park, Sang-June(박 상 준)*****
Hyun, Yong J. (현 용 진)*****

A new chi-square test is proposed to assess significance of attributes for nonparametric conjoint models. The key idea is to form subsets of rankings and test the dependence between the attribute levels and the sets of rankings. The null hypothesis states that the rankings for profiles with the focal attribute are distributed randomly among the sets of rankings. The approach is simple, easy to use, and can be applied at the individual level as well as at the aggregate level. It can be used for the trade-off approach as well as for the full profile approach.

Key words: Nonparametric, Conjoint Analysis, Chi-square test, Trade-off Approach

* Professor of Marketing, KAIST Business School, 207-43, Chongryang-2, Dongdaemun-ku, Seoul, Korea.
(mhhahn@kgsn.kaist.ac.kr)

** Montgomery Ward Professor of Marketing, Department of Marketing, Kellogg Graduate School of Management, Northwestern University(laksh@kellogg.northwestern.edu)

*** Doctoral Student, KAIST Business School(lilac@kgsn.kaist.ac.kr)

**** President, DMS Lab, Inc., Seoul, Korea(jshyun@dms-lab.com)

***** Associate Professor of School of Business Administration, Chonbuk National University(psj@chonbuk.ac.kr)

***** Professor of Marketing, KAIST Business School(yhyun@kgsn.kaist.ac.kr)

I. Introduction

Conjoint Analysis is widely used to measure the utility that consumers place on product and service attributes. The methodology has been applied to diverse situations including competitive analysis, pricing, market segmentation, positioning, advertising, and distribution (Cattin and Wittink 1982; Green and Srinivasan 1978, 1990; Louviere, Hensher, and Swait 2000; Wittink and Cattin 1989).

For many of the applications, it is important to test the significance of attribute levels. Kohli (1988) summarizes the importance of assessing attribute significance as follows. (1) It allows testing whether two segments differ in terms of the benefits they seek from the product or service offering. (2) It helps in eliminating insignificant attributes in the choice simulation, thereby increasing the efficiency of the simulation. (3) It reduces the number of attributes or allows the inclusion of other important attributes in subsequent conjoint studies for the same product class. Furthermore, testing the significance of attribute levels at the individual level is crucial for benefit segmentation.

In conjoint analysis, various approaches for data collection and data analysis have been proposed. When rating data are collected,

regression analysis is the most widely used method to estimate part worth utilities. A t-test can be employed to test significance of attributes at the individual level as well as at the aggregate level. When ranking data are collected, MONANOVA, ordinal logit or LINMAP may be used to estimate the part worth utilities.¹⁾ If ordinal logit is applied, the Wald chi-square test can be employed to test attribute significance. Ordinal logit, however, is typically appropriate for estimating the utilities at the aggregate level and not at the individual level.

MONANOVA can be applied for both individual and aggregate level data. When ordinal data are analyzed by MONANOVA, the importance weight is most widely used to assess the meaningfulness of attributes even though it does not have any statistical implications. Kohli(1988) has shown that Kruskal-Wallis test (Kruskal and Wallis 1952) and Friedman test (Friedman 1937) can be applied for testing attribute significance at the aggregate level when the full-profile approach is adopted. Unlike the Friedman test, the Kruskal-Wallis test can be applied at the individual level also.

The purpose of this paper is to propose a new nonparametric test of attribute significance when ordinal data are collected at the individual level. The new method is simple and

1) In practice, regression analysis is also used with ranking data.

easy to use. It can be applied with individual level data as well as aggregate data. It can be used for the trade-off approach as well as the full-profile approach. Furthermore, it has the advantage that researchers can make the test more conservative or less conservative depending on the testing situation.

II. METHOD

2.1 A simple illustration

In his article, Kohli(1988, p.129) used the hypothetical data shown in <Table 1> to demonstrate his test method. We use the same example to illustrate how our proposed test is applied.

Let us focus on attribute 1 of person 1. We group the rankings into 2 subsets: a subset with ranks 1 and 2 as its elements and the other with ranks 3 and 4 as the elements, i.e., {1,2} and {3,4}. The former set has the high

ranks and the latter set has the low ranks as its elements. Note that the rankings for profiles with level 1 of attribute 1 are 1 and 2. Thus, the number of ranks in set {1,2} is 2 whereas that of ranks in set {3,4} is 0. On the other hand, the rankings for profiles with level 2 of attribute 1 are 3 and 4. Thus, the number of ranks in set {1,2} is 0 while that of ranks in set {3,4} is 2.

Simple logic tells us that attribute 1 should be significant for person 1 because the attribute levels are strongly associated with the ranks of the profiles. <Table 2> shows the cross-tabulation analysis for this case. The null hypothesis is that the attribute levels and rank sets are independent. Applying the chi-square test, we find that $\chi^2 = 4$. It is greater than 3.84, the criterion value for 1 degree of freedom at $\alpha = .05$, rejecting the null hypothesis.

For person 2, level 2 of attribute 1 is associated with high ranks while level 1 is associated with low ranks. A similar chi-square test shows that attribute 1 is also significant for person 2. However, as we can expect from

<Table 1> Rankings of Kohli's Hypothetical Data

PROFILE	ATTRIBUTE		RANKING	
	1	2	Person 1	Person 2
1	LEVEL 1	LEVEL 1	1	4
2	LEVEL 1	LEVEL 2	2	3
3	LEVEL 2	LEVEL 1	4	1
4	LEVEL 2	LEVEL 2	3	2

the heterogeneity between the two individuals, the aggregate level chi-square is not significant ($\chi^2 = 0$). Applying the same test, we can show that attribute 2 is not significant either for individuals or for the group. This is expected from the data as the attribute levels are not associated with high or low rankings.

2.2 A notational presentation of the test

Let n denote the number of profiles evaluated by each respondent. For notational convenience, we assume that '1' denotes the rank of the most preferred profile and 'n' that of the least preferred profile.

We define Q disjoint subsets of ranks a priori. Let S_q ($q=1,2,\dots,Q$) denote the disjoint subsets of ranks.

$$S_q = \left\{ \sum_{i=0}^{q-1} a_i + 1, \sum_{i=0}^{q-1} a_i + 2, \dots, \sum_{i=0}^{q-1} a_i + a_q \right\}$$

where $q=1,2,\dots,Q$, $a_0=0$, and $\sum_{i=1}^Q a_i = n$

Consider an attribute with m levels. Let us

assume that each level, denoted as j , appears in n_j profiles. Let $X_{j,k}$, $k=1,2,\dots,n_j$, denote the n_j profiles in which level j of the attribute appears. For example, consider an attribute with three levels ($m=3$). Let two levels ($j=1, 2$) appear in two profiles ($n_1 = n_2 = 2$) and one level ($j=3$) appear in four profiles ($n_3 = 4$; $n = 8$). Then $X_{1,1}, X_{1,2}, X_{2,1}, X_{2,2}$ denote the four profiles in which level 1 and 2 appear and $X_{3,1}, X_{3,2}, X_{3,3}, X_{3,4}$ denote the four profiles in which level 3 appears. Additionally, let $R(X_{j,k})$ be the rank of profile $X_{j,k}$ and $f_{jq} = \sum_{k=1}^{n_j} I(R(X_{j,k}))$

$$\text{where } I(R(X_{ik})) = \begin{cases} 1, & \text{if } R(X_{ik}) \in S_q \\ 0, & \text{otherwise} \end{cases}$$

Then, f_{jq} represents the number of profiles with level j of the focal attribute in the subset S_q . The f_{jq} 's can be arranged in the following contingency table of F where the rows represent the levels of the focal attribute and the columns are the subsets of ranks:

(Table 2) Cross Tabulation of Attribute Levels and Rank Sets for Attribute 1

Rank Sets	Person 1		Person 2		Total	
	{1,2}	{3,4}	{1,2}	{3,4}	{1,2}	{3,4}
Level 1	2	0	0	2	2	2
Level 2	0	2	2	0	2	2

$$F = \left\{ \begin{array}{cccc} f_{11} & . & . & f_{1Q} \\ . & . & . & . \\ . & . & f_{jQ} & . \\ . & . & . & . \\ f_{m1} & . & . & f_{mQ} \end{array} \right\}$$

We apply the well known chi-square test of independence between two nonmetric variables using the contingency matrix F . In our case, we are applying the chi-square test of independence between levels of an attribute and subsets of ranks. The null hypothesis states that the attribute levels and the rank sets are independent.

$$H_0 : P_{jq} = P_j P_q, \quad j=1,2,\dots,m \text{ and } q=1,2,\dots,Q$$

where P_{jq} is the probability that level j appears in the rank set q of the population, and P_j and P_q are marginal probabilities of level j and the rank set q , respectively, appearing in the population.

The test statistic h is simply the chi-square computed as:

$$h = \sum_{j=1}^m \sum_{q=1}^Q \frac{(f_{jq} - E[f_{jq}])^2}{E[f_{jq}]} \sim \chi^2((Q-1) \times (m-1))$$

III. Comparison with Kohli's Test

Noting the importance of assessing significance

of attributes, Kohli(1988) proposed nonparametric tests of attribute significance based on the Kruskal and Wallis(1952) test and the Friedman(1937) test. We will focus on comparing our test with the Kruskal-Wallis test as it can be applied with individual level data. We will denote the statistic developed by Kohli as c_i .

The null hypothesis in Kohli's test is a random association hypothesis such that each of the product profiles is assigned any one of the n ranks with probability $1/n$. Under the null hypothesis, expectation and variance of r_{ij} are calculated where r_{ij} is the sum of ranks for attribute level j for respondent i . The distribution of r_{ij} asymptotically approaches the normal distribution. Thus, the square of the standardized r_{ij} has an asymptotic chi square distribution with 1 degree of freedom. The test statistic c_i is the sum of c_{ij} over attribute levels after a correction for the dependence of rankings as follows:

$$\begin{aligned} c_i &= \sum_{j=1}^m \{(n - n_j) \div n\} c_{ij} \\ &= \frac{12}{n(n+1)} \sum_{j=1}^m \{r_{ij} - (1/2)n_j(n+1)\}^2 \div n_j \end{aligned}$$

$$\text{where } c_{ij} = \{r_{ij} - E(r_{ij})\}^2 \div \text{var}(r_{ij}),$$

and n_j is the number of profiles in which attribute level j appears.

The c_i has a limiting chi square distribution with $(m-1)$ degrees of freedom. When the test

is applied to a segment, the c_i is summed across respondents $i=1, \dots, I$. Because each c_i has $(m-1)$ degrees of freedom, the sum c has $I(m-1)$ degrees of freedom.

Let us denote our newly proposed statistic for individual i by h_i , where $i=1, \dots, I$. The h_i and c_i are based on a similar logic, i.e., if an attribute is significant, ranks associated with the attribute should not be distributed randomly. Both statistics are based on the random-association null hypothesis.

Although the two tests rely on the same null hypothesis, the statistics testing the hypothesis are different. Kohli's statistic is based on the sum of ranks assigned to profiles that include a certain attribute level. If the actual sum of ranks is much different from the expected sum of ranks, then the null hypothesis is likely to be false.

On the other hand, to apply our method, researchers should predefine sets of ranks. The number of times a particular attribute level j occurs in the set of ranks is counted. If the distribution of actual number in the sets of ranks is much different from that of the expected number, then the null hypothesis is likely to be false.

IV. Defining the Number of Subsets of Ranks

The new method requires researchers to

define subsets of ranks a priori. Defining subsets has both advantages and disadvantages. One disadvantage is in determining an appropriate number of subsets. A natural starting point is to define two subsets of ranks having equal number of elements, one with high ranks and the other with low ranks. The number of subsets can be increased to identify the most adequate test. The numerical analysis illustrated in <Table 3> can be used to find the appropriate number of groups for testing.

<Table 3> illustrates the effects of using different numbers of subsets when there are attributes with 2 levels and 3 levels. In <Table 3-(a) and 3-(b)>, we assumed that ranks were obtained using 8 profiles, while in <Table 3-(c) and 3-(d)>, we assumed that ranks were obtained using 12 profiles. Two level attribute cases are reported in <Table 3-(a) and 3-(c)> while three level cases are in <Table 3-(b) and (d)>. For each table, h statistics are calculated when the ranks were grouped into two, three, and four disjoint subsets, respectively. Also, for comparison, Kohli's c statistic is also reported. Various hypothetical cases were examined to see the effects of different groupings. Nine cases are presented in <Table 3-(a), (b), and (c)> whereas ten cases are presented in <Table 3-(d)>. The tables also present the criterion chi-square values for $\alpha=.10$.

<Table 3-(a)> shows how grouping of ranks affects the test outcomes for 2 level attributes when 8 profiles are used. In the table, the

<Table 3> The "h" and "c" statistic for various cases
 Table 3-(a): Two level attribute with 8 profiles

Ranks		Effects of Grouping								c-values			
		Two Groups		Three Groups		Four Groups							
Cases	{1,2,3,4}	{5,6,7,8}	{1,2,3}	{4,5}	{6,7,8}	h-values	{1,2}	{3,4}	{5,6}	{7,8}	h-values	c-values	
Case 1	Level 1	4	0	3	1	0	6*	2	2	0	0	8*	5.33*
	Level 2	0	4	0	1	3		0	0	2	2		
Case 2	Level 1	3	1	3	1	0	6*	2	1	1	0	4	4.08*
	Level 2	1	3	0	1	3		0	1	1	2		
Case 3	Level 1	3	1	3	0	1	5.33*	2	1	0	1	4	1.33
	Level 2	1	3	0	2	2		0	1	2	1		
Case 4	Level 1	3	1	2	2	0	5.33*	2	1	1	0	4	3*
	Level 2	1	3	1	0	3		0	1	1	2		
Case 5	Level 1	3	1	2	1	1	0.67	2	1	1	0	4	2.08
	Level 2	1	3	1	1	2		0	1	1	2		
Case 6	Level 1	3	1	2	1	1	0.67	1	2	0	1	4	0.08
	Level 2	1	3	1	1	2		1	0	2	1		
Case 7	Level 1	2	2	2	1	1	0.67	2	0	2	0	8*	1.33
	Level 2	2	2	1	1	2		0	2	0	2		
Case 8	Level 1	2	2	2	0	2	2.67	2	0	0	2	8*	0
	Level 2	2	2	1	2	1		0	2	2	0		
Case 9	Level 1	2	2	2	1	1	0.67	1	1	1	1	0	0.33
	Level 2	2	2	1	1	2		1	1	1	1		
Criterion value ($\alpha = 0.10$)							4.61					6.25	2.71

* Significant at $\alpha = 0.10$

<Table 3> The "h" and "c" statistic for various cases
 Table 3-(b) : Three level attribute with 8 profiles

Ranks		Effects of Grouping								c-values		
		Two Groups		Three Groups		Four Groups		h-values	h-values			
Cases	Level	{1,2,3,4}	{5,6,7,8}	{1,2,3}	{4,5}	{6,7,8}	{1,2}			{3,4}	{5,6}	{7,8}
Case 1	Level 1	2	0	2	0	0	2	0	0	0	0	6*
	Level 2	2	2	1	2	1	0	2	2	2	0	
	Level 3	0	2	0	0	2	0	0	0	2	0	
Case 2	Level 1	2	0	2	0	0	2	1	1	0	0	5.73*
	Level 2	2	2	1	2	1	1	1	2	2	0	
	Level 3	0	2	0	0	2	0	0	0	2	0	
Case 3	Level 1	2	0	2	2	1	1	1	1	0	0	4.17
	Level 2	2	2	1	2	1	1	1	1	1	1	
	Level 3	0	2	0	0	2	0	0	0	2	0	
Case 4	Level 1	2	0	2	1	2	1	1	1	1	1	4.5
	Level 2	2	2	1	1	0	0	1	1	0	0	
	Level 3	0	2	0	0	2	1	1	2	2	0	
Case 5	Level 1	1	1	1	1	0	0	0	0	0	0	4.13
	Level 2	3	1	2	1	1	1	2	1	1	0	
	Level 3	0	2	0	0	2	0	0	0	2	0	
Case 6	Level 1	2	0	1	1	0	1	1	1	0	0	2.67
	Level 2	2	2	2	0	2	1	1	1	1	1	
	Level 3	0	2	0	1	1	0	0	0	1	1	
Case 7	Level 1	1	1	1	1	0	1	1	1	0	0	1.5
	Level 2	2	2	2	0	2	1	1	1	1	1	
	Level 3	1	1	0	1	1	0	1	1	0	0	
Case 8	Level 1	1	0	1	0	1	0	1	0	1	1	0.17
	Level 2	2	2	1	2	1	1	0	2	2	0	
	Level 3	1	1	1	0	1	1	0	0	1	1	
Case 9	Level 1	1	1	1	0	1	1	1	0	0	1	0.33
	Level 2	2	2	1	1	0	1	1	0	0	1	
	Level 3	1	1	1	1	2	1	1	1	1	1	
Criterion value ($\alpha = 0.10$)												10.64
												4.61

* Significant at $\alpha = 0.10$

<Table 3> The "h" and "c" statistic for various cases
 Table 3-(c): Two level attribute with 12 profiles

Ranks	Effects of Grouping														
	Two Groups			Three Groups			Four Groups			h-values	c-values				
	{1,2,3,4,5,6}	{7,8,9,10,11,12}	h-values	{1,2,3,4}	{5,6,7,8}	{9,10,11,12}	h-values	{1,2,3}	{4,5,6}			{7,8,9}	{10,11,12}		
Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9							
Level 1	[1,2,3,4,5,6]	[1,2,3,4,5,9]	[1,2,3,4,7,8]	[1,2,3,7,8,9]	[2,3,4,5,10,11]	[2,4,5,6,10,11]	[1,2,3,10,11,12]	[1,3,5,7,9,11]	Level 2	[7,8,9,10,11,12]	[6,7,8,10,11,12]	[3,8,9,10,11,12]	[5,6,10,11,12]	[4,5,6,7,8,9]	[2,4,6,8,10,12]
Case 1	Level 1	6	0	12*	4	2	0	8*	3	3	0	0	12*	8.31*	
Case 1	Level 2	0	6		0	2	4		0	0	3	3			
Case 2	Level 1	5	1	5.33*	4	1	1	6*	3	2	1	0	6.67*	5.77*	
Case 2	Level 2	1	5		0	3	3		0	1	2	3			
Case 3	Level 1	5	1	5.33*	3	3	0	6*	2	3	1	0	6.67*	5.03*	
Case 3	Level 2	1	5		1	1	4		1	0	2	3			
Case 4	Level 1	4	2	1.33	4	2	0	8*	3	1	2	0	6.67*	5.03*	
Case 4	Level 2	2	4		0	2	4		0	2	1	3			
Case 5	Level 1	3	3	0	3	2	1	2	3	0	3	0	12*	2.08	
Case 5	Level 2	3	3		1	2	3		0	3	0	3			
Case 6	Level 1	4	2	1.33	3	1	2	2	2	2	0	2	4	0.41	
Case 6	Level 2	2	4		1	3	2		1	1	3	1			
Case 7	Level 1	4	2	1.33	2	2	2	0	1	3	0	2	6.67*	0.03	
Case 7	Level 2	2	4		2	2	2		2	0	3	1			
Case 8	Level 1	3	3	0	3	0	3	6*	3	0	0	3	12*	0	
Case 8	Level 2	3	3		1	4	1		0	3	3	0			
Case 9	Level 1	3	3	0	2	2	2	0	2	1	2	1	1.33	0.23	
Case 9	Level 2	3	3		2	2	2		1	2	1	2			
Criterion value ($\alpha = 0.10$)				2.71				4.61					6.25	2.71	

* Significant at $\alpha = 0.10$

<Table 3> The "h" and "c" statistic for various cases
 Table 3-(d): Three level attribute with 12 profiles

Ranks	Effects of Grouping												c-values
	Two Groups			Three Groups			Four Groups			h-values	h-values		
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9			Case 10	
Level 1	[1,2,3,4]	[1,2,4,5]	[1,2,3,11]	[1,2,4,5]	[1,2,5,6]	[1,3,4,5]	[1,2,3,8]	[1,3,5,6]	[1,2,11,12]	[1,4,7,10]	9.85*		
Level 2	[5,6,7,8]	[3,6,7,9]	[5,6,7,8]	[6,7,8,12]	[3,7,8,10]	[2,6,7,10]	[4,5,6,12]	[2,7,8,9]	[5,6,7,8]	[2,5,8,11]	8.12*		
Level 3	[9,10,11,12]	[8,10,11,12]	[4,9,10,12]	[3,9,10,11]	[4,9,11,12]	[8,9,11,12]	[7,9,10,11]	[4,10,11,12]	[3,4,9,10]	[3,6,9,12]	3.12		
Case 1	Level 1	4	0	8*	4	0	0	24*	3	1	0	16*	9.85*
Case 1	Level 2	2	2	0	0	0	0	0	0	2	2	0	0
Case 1	Level 3	0	4	0	0	4	0	0	0	0	1	3	0
Case 2	Level 1	4	0	8*	3	1	0	7.5	2	2	0	12*	8.12*
Case 2	Level 2	2	2	2	1	2	1	1	1	1	2	0	0
Case 2	Level 3	0	4	0	0	1	3	15*	0	0	1	3	0
Case 3	Level 1	3	1	2	3	0	1	0	3	0	0	1	12*
Case 3	Level 2	2	2	2	0	4	0	0	0	2	2	0	0
Case 3	Level 3	1	3	0	1	0	3	10.5*	0	1	1	2	8
Case 4	Level 1	4	0	6*	3	1	0	0	2	2	0	0	5.65*
Case 4	Level 2	1	3	3	0	3	1	1	0	1	2	1	0
Case 4	Level 3	1	3	3	1	0	3	6	1	0	1	2	4.77*
Case 5	Level 1	4	0	6*	2	2	0	0	2	2	0	0	8
Case 5	Level 2	1	3	3	1	2	1	7.5	1	0	2	1	7.04*
Case 5	Level 3	1	3	3	1	0	3	7.5	0	1	1	2	0
Case 6	Level 1	4	0	8*	3	1	0	0	2	2	0	0	8
Case 6	Level 2	2	2	2	1	2	1	1	1	1	1	1	0
Case 6	Level 3	0	4	0	0	1	3	7.5	0	0	2	2	5.12*
Case 7	Level 1	3	1	6*	3	1	0	0	3	0	1	0	16*
Case 7	Level 2	3	1	3	1	2	1	0	0	3	0	1	0
Case 7	Level 3	0	4	0	0	1	3	6	2	2	0	0	16*
Case 8	Level 1	4	0	6*	2	2	0	0	2	2	0	0	4.65*
Case 8	Level 2	1	3	3	1	2	1	1	1	0	3	0	0
Case 8	Level 3	1	3	3	1	0	3	12*	0	1	0	3	0
Case 9	Level 1	2	2	0	2	0	2	1.5	2	0	0	2	8
Case 9	Level 2	2	2	2	0	4	0	0	0	2	2	0	0
Case 9	Level 3	2	2	2	2	0	2	1	1	1	1	1	0
Case 10	Level 1	2	2	0	2	1	1	1.5	1	1	1	1	0
Case 10	Level 2	2	2	2	1	2	1	1	1	1	1	1	0
Case 10	Level 3	2	2	2	1	1	2	7.78	1	1	1	1	10.64
Criterion value ($\alpha = 0.10$)				4.61				7.78					4.61

* Significant at $\alpha = 0.10$

attribute level is perfectly associated with ranks in Case 1. In Cases 2, 3, and 4, the attribute level appears to be associated with the ranks but not as perfectly as Case 1. In Cases 5, 6, and 7, the association between attribute levels and ranks is ambiguous. In Case 8, the attribute level 1 is associated with either very high ranks or very low ranks. In this case, level 1 is either highly liked or highly disliked. Finally, in Case 9, the ranks are almost randomly assigned across attribute levels. <Table 4> shows sets of profiles and rankings that correspond to Cases illustrated in <Table 3-(a)>.

Kohli's statistic finds the attribute significant in Cases 1, 2, and 4. For example, the statistic value is 5.33 for Case 1. The value is greater than the criterion value 2.71. Because Kohli's

statistic is based on the difference between mean ranks of attribute levels, it is expected that the attribute is significant for Cases with the largest differences between mean ranks of level 1 and 2.

The new test statistic h finds the attribute significant in Cases 1, 2, 3, and 4 when three groups are used. For example, consider Case 1 when three groups are used. There are three ranks in the rank set {1,2,3}, and one in the set {4,5} for the attribute level 1. For level 2, there is one rank in the set {4,5} and three in the set {6,7,8}. From the data, the h -statistic value is calculated as 6. It exceeds the criterion value 4.61. However, when two groups are used, the h -test is not satisfactory in finding attribute significance as the attribute is significant in only the perfect case, i.e., Case 1.

<Table 4> A Full Profile Set and Rankings Corresponding to Cases in Table 3-(a)

Profile	Attribute			Case								
	1	2	3	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	2	1	1	1
2	1	2	1	2	2	2	2	2	3	2	2	2
3	2	1	1	5	4	4	3	3	1	3	5	5
4	2	2	1	6	6	5	6	5	5	4	6	6
5	1	1	2	3	3	3	4	4	4	5	7	7
6	1	2	2	4	5	8	5	6	8	6	8	8
7	2	1	2	7	7	6	7	7	6	7	3	3
8	2	2	2	8	8	7	8	8	7	8	4	4
(a) Mean rank of level 1				2.5	2.75	3.5	3	3.25	4.25	3.5	4.5	4
(b) Mean rank of level 2				6.5	6.25	5.5	6	5.75	4.75	5.5	4.5	5
Difference between (b) and (a)				4	3.5	2	3	2.5	0.5	2	0	1

It appears that making four groups is not appropriate either as the attribute is not significant for Cases 2, 3, and 4. Four groups may be too many when we have only 8 profiles. However, it is remarkable that our test finds the attribute significant in Cases 7 and 8 when we have four groups. These are cases where there may be interactions between attributes. For example, consider Case 8 in which level 1 is either highly liked or highly disliked. Observing <Table 4>, we find that attribute level 1 is liked very much when it is combined with level 1 of attribute 3 and disliked very much when it is combined with level 2 of attribute 3. This strongly suggests that attribute 1 and attribute 3 have an interaction effect on consumer preference. Similarly, Case 7 shows a potential interaction. In this case, level 1 of attribute 3 is preferred to level 2 of attribute 3. Furthermore, for each level of attribute 3, level 1 of attribute 1 is preferred to level 2 of attribute 1. It could be helpful to increase the number of groups for the new test when researchers need to find interactions among attributes.

<Table 3-(b)> shows cases for three level attributes when data are collected with 8 profiles. Again, attribute levels are perfectly associated with ranks in Case 1. Cases 2, 3, and 4 are ones where the attribute levels appear to have some association to the ranks. The association is ambiguous in Cases 5, 6, and 7. In Case 8, attribute level 1 and 3 are either

highly liked or highly disliked. The ranks are almost randomly assigned to the attribute levels in Case 9.

Kohli's test finds the attribute significant for Cases 1, and 2. When three groups are used, the new test finds the attribute significant for Cases 1, 2, and 3. The test becomes less than satisfactory when two groups or four groups are used. However, we can still apply two groups or four groups if we are willing to adopt more generous α levels for the test. By slightly expanding the critical region, i.e., for $\alpha = 0.15$, Kohli test shows significance for Case 1, 2, 3, 4, and 5, and the h-test shows significance for Case 1, 2, 3, 4, and 6 for two groups, and Case 1, 2, 4, and 5 for four groups.

<Table 3-(c)> shows 2 level attribute cases with 12 profiles. Compared to <Table 3-(a)>, with the increased number of data, we find that our test becomes more sensitive not only for three groups, but also for two or four groups. The same is true with Kohli's test. Also, notable is that interaction effects are much better captured with the increased number of data. For example, in Case 8, we find that level 1 is associated with either very preferred profiles or very disliked profiles depending on levels of the other attributes. Thus, it is highly likely that there is an interaction effect between attribute 1 and attribute 2 or attribute 3.

<Table 3-(d)> shows 3 level attribute cases with 12 profiles. Comparing with <Table

3-(b)), we also find that the tests become more sensitive with the increased number of data. Kohli's test becomes almost too sensitive in that it finds the attribute significant in all the cases shown except for Cases 3, 9, and 10. The h -test is also too sensitive when we have two groups. It is more appropriate to use three or four groups. Again, it is interesting that the new test can capture interaction effects when we use three or four groups. For example, in Case 9 with three groups, we see that level 1 and level 3 are highly liked or highly disliked depending on levels of the other attributes in the profiles.

The illustration in <Table 3> shows that researchers can find an appropriate number of groups for the h -test by doing a relatively simple analysis. It also shows that when there is a small number of profiles, it is not desirable to have too many groups. With increased number of data, researchers may typically choose between three or four groups. If they are interested in finding interactions, it may be better to have more groups.

V. Further Applications

5.1 Testing attribute significance at aggregate level

It is straightforward to extend the h -test

from the individual level to the aggregate level. We simply aggregate the F matrices over individuals in a group to get an aggregate level F matrix. The same chi-square test can be applied to the aggregate F matrix.

5.2 Applying the h -test for Trade-off Approach

Although not as popular as the full-profile approach nowadays, trade-off matrices are still used to collect data for nonparametric Conjoint Analysis (e.g., Kivetz, Netzer, and Srinivasan 2004). Currently, there is no statistical method to test significance of attributes when the tradeoff design is employed. The new method can serve this purpose. Here, we illustrate an application of the h -test to the data reported by Johnson(1974, p.122) for passenger cars. The mathematical description of the method for the tradeoff approach is briefly presented in the Appendix. Considered attributes are price, seating capacity, top speed, and months of warranty. <Table 5> shows the rankings provided by one individual.

Suppose our focal attribute is price. Unlike the full profile approach where there is only one set of ranks, there are multiple sets of ranks for the tradeoff approach. In the example, there are 3 sets of ranks for price, one for each of other tradeoff attributes. Suppose we consider three groups of ranks such as {1,2,3}, {4,5,6}, and {7,8,9}. Let us

<Table 5> Rankings of Trade-off Matrices Reported in Johnson's (1974) Study

		TOP SPEED (MPH)			SEATING CAPACITY			MONTHS OF WARRANTY		
		130	100	70	2	4	6	60	12	3
PRICE	\$2,500	1	2	5	2	1	3	1	3	4
	\$4,000	3	4	6	6	5	4	2	5	6
	\$6,000	7	8	9	8	7	9	7	8	9
TOP SPEED	130				2	1	3	1	2	5
	100				5	4	6	3	4	6
	70				8	7	9	7	8	9
SEATING CAPACITY	2							2	5	8
	4							1	4	7
	6							3	6	9

compute a contingency matrix when price is evaluated against top speed. For profiles with price of \$2,500, we know that there are two ranks belonging to the subset {1,2,3}, one to {4,5,6}, and none to {7,8,9}. For profiles with price of \$4,000, there is one rank in the subset {1,2,3}, two in {4,5,6}, and none in {7,8,9}. Also, for profiles with price of \$6,000, all the ranks are in the subset {7,8,9}. Let us denote the contingency matrix as F^1 . Then,

$$F^1 = \begin{Bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{Bmatrix}$$

In the same way, we can find the contingency matrix when price is evaluated against seating capacity F^2 , and that when price is evaluated against months of warranty

F^3 . The contingency table F is obtained by adding the three contingency matrices resulting in

$$F = \begin{Bmatrix} 7 & 2 & 0 \\ 2 & 7 & 0 \\ 0 & 0 & 9 \end{Bmatrix} .$$

If the levels of price and subsets of ranks are independent, we would expect that every entry of the F matrix to be the number 3. Therefore, the h -test statistic is computed as:

$$\begin{aligned} x^2 &= (1/3)[(7-3)^2+(2-3)^2+(0-3)^2+(2-3)^2 \\ &\quad +(7-3)^2+(0-3)^2+(0-3)^2+(0-3)^2 \\ &\quad +(9-3)^2]=35.33. \end{aligned}$$

Since it is greater than the criterion $x^2(4) = 7.78$ for $\alpha = 0.1$, we conclude that price is

an important attribute. Similarly, top speed ($x^2 = 25.33$) and warranty ($x^2 = 15.33$) are significant, whereas seating capacity ($x^2 = 0$) is not.

For the tradeoff approach, our experience suggests that it is usually appropriate to set the number of subsets of ranks equal to the number of levels of the focal attribute. It is always feasible. In addition, it is usually the most conservative test. However, researchers may do a similar analysis to that introduced for the full profile approach to find an appropriate number of groups of ranks.

VI. An Empirical Comparison

Our objective in this section is to illustrate how the new test is applied with real data and to empirically compare the test with other methods. For the empirical comparison, we designed a dormitory study as shown in <Table 6>. We use three two-level attributes, i.e., number of roommates, distance to classrooms, and whether undergraduates are allowed in the

dormitory or not, and one three-level attribute, i.e., monthly rent. The profiles used as stimuli are shown in <Table 7>. Note that orthogonal fractional factorial method was used in generating estimation profiles so that the attributes are orthogonal. We have fourteen profiles in total, eight are used for estimation and the remaining six for validation.

One hundred and thirteen subjects, who are MBA students in a university located in Asia, responded by either ranking the fourteen profiles or rating them using a 9-point scale. We will mainly discuss the results of analyzing the ranking data for which our test is most appropriate. Later, we will briefly comment on the results applied to the rating data.

Among the subjects, 60 responded by ranking the profiles. We estimated the parameters using MONANOVA, ordered logit, and regression analysis. Based on the estimated models, ranks of holdout profiles were predicted for each respondent. Responses showing very low average correlation (lower than .4) between predicted and actual ranks over the three estimation models were eliminated from further analysis leaving 43 cases.

<Table 6> Design of the Dormitory Study

	RENT	NUMBER OF ROOMMATES	DISTANCE TO CLASS	UNDERGRADUATE ALLOWED
LEVEL 1	\$210	NONE	10 MINUTES	NO
LEVEL 2	\$315	ONE	20 MINUTES	YES
LEVEL 3	\$420	-	-	-

<Table 7> Stimuli for the Dormitory Study

PROFILES	RENT	NUMBER OF ROOMMATES	DISTANCE TO CLASS	UNDERGRADUATE ALLOWED
1	\$210	NONE	10 MINUTES	YES
2	\$210	ONE	20 MINUTES	NO
3	\$315	NONE	20 MINUTES	NO
4	\$315	NONE	10 MINUTES	YES
5	\$420	NONE	10 MINUTES	NO
6	\$420	ONE	20 MINUTES	YES
7	\$315	ONE	20 MINUTES	YES
8	\$315	ONE	10 MINUTES	NO
<u>Holdout profiles</u>				
PROFILES	RENT	NUMBER OF ROOMMATES	DISTANCE TO CLASS	UNDERGRADUATE ALLOWED
1	\$420	NONE	10 MINUTES	NO
2	\$210	ONE	20 MINUTES	YES
3	\$315	NONE	20 MINUTES	YES
4	\$315	ONE	10 MINUTES	NO
5	\$420	ONE	10 MINUTES	YES
6	\$210	NONE	20 MINUTES	NO

6.1 Comparison of face validity

For the remaining 43 cases, we applied the new chi-square test as well as Kohli's test for each individual case estimated by the MONANOVA model. The Wald chi-square test was applied when parameters were estimated by the ordered logit model. For regression analysis, t-tests for the three two-level attributes and F-tests for the one three-level attribute were used even though the ranks are not interval scaled. <Table 8>

shows significant attributes for each test applied for each respondent. In the table, there are two columns of results for the *h*-test. The left column shows all the attributes significant in the *h*-test. The right column shows the attributes with significant main effects, excluding attributes that reveal interactions. Possible interactions were observed in Cases 5, 6, 8, 9, 16, 20, and 36.

To check the face validity of the tests, Table 8 includes a column that shows "likely-to-be-significant" attributes identified from the

<Table 8> Significant Attributes for Various Tests with Ranking Data

Respondent	Significant attributes *					Likely-to-be-Significant
	h-test (1)	h-test (2) **	Kohli's test	Regression (t-test)	Ordered Logit (Wald χ^2 test)	
1	RENT	RENT	NONE	NONE	NONE	RENT
2	MATE	MATE	MATE	MATE, RENT	NONE	MATE
3	MATE	MATE	MATE	MATE, UNDER	NONE	MATE
4	NONE	NONE	NONE	NONE	NONE	RENT
5	MATE, DISTANCE	MATE	MATE	MATE	NONE	MATE
6	MATE	NONE	NONE	RENT	NONE	RENT
7	MATE	MATE	MATE	MATE, RENT	NONE	MATE
8	MATE	NONE	NONE	MATE, RENT	NONE	RENT
9	MATE	NONE	NONE	MATE, UNDER, RENT	NONE	RENT
10	RENT	RENT	RENT	RENT	NONE	RENT
11	MATE	MATE	MATE	MATE, DISTANCE, UNDER, RENT	UNDER	MATE
12	MATE, DISTANCE	MATE, DISTANCE	MATE, DISTANCE	MATE, DISTANCE, RENT	NONE	MATE, DISTANCE
13	MATE	MATE	MATE	MATE, RENT	NONE	MATE
14	RENT	RENT	RENT	RENT	NONE	RENT
15	RENT	RENT	RENT	MATE, DISTANCE, RENT	NONE	RENT
16	MATE	NONE	NONE	MATE, RENT	NONE	RENT
17	MATE	MATE	MATE	MATE, RENT	NONE	MATE
18	UNDER	UNDER	UNDER	NONE	UNDER	UNDER
19	MATE	MATE	NONE	MATE, RENT	MATE, RENT	RENT
20	UNDER	NONE	NONE	RENT	UNDER, RENT	RENT
21	RENT	RENT	NONE	NONE	NONE	RENT
22	MATE	MATE	MATE	MATE, RENT	NONE	MATE, RENT
23	MATE	MATE	MATE	MATE, RENT	NONE	MATE
24	MATE, DISTANCE	MATE, DISTANCE	MATE	MATE	RENT	MATE
25	MATE	MATE	MATE	MATE, RENT	NONE	MATE, RENT
26	RENT	RENT	RENT	MATE, RENT	NONE	RENT
27	MATE	MATE	NONE	NONE	DISTANCE	NONE
28	MATE, DISTANCE	MATE, DISTANCE	MATE, DISTANCE	MATE, DISTANCE	NONE	MATE, DISTANCE
29	MATE	MATE	MATE	MATE, DISTANCE, UNDER, RENT	NONE	MATE, RENT
30	UNDER	UNDER	UNDER	MATE, UNDER	RENT	UNDER
31	MATE	MATE	MATE	MATE, RENT	NONE	MATE, RENT
32	RENT	RENT	RENT	RENT	NONE	RENT
33	MATE	MATE	MATE	NONE	MATE, RENT	MATE, RENT
34	MATE	MATE	MATE	MATE	RENT	MATE
35	RENT	RENT	RENT	RENT	NONE	RENT
36	MATE, UNDER	UNDER	UNDER	RENT	MATE, UNDER	UNDER
37	RENT	RENT	RENT	MATE, RENT	NONE	RENT
38	MATE	MATE	MATE	MATE, RENT	NONE	MATE, RENT
39	NONE	NONE	NONE	NONE	NONE	NONE
40	RENT	RENT	RENT	RENT	NONE	RENT
41	MATE	MATE	MATE	MATE, RENT	NONE	MATE
42	RENT	RENT	RENT	RENT	NONE	RENT
43	MATE	MATE	MATE	MATE, DISTANCE, UNDER, RENT	NONE	MATE, RENT

* Significant at $\alpha = 0.10$

** Significant attributes after eliminating those with interaction effects

*** MATE: Number of roommates, DISTANCE: Distance to class, UNDER: Undergraduate allowed

raw data. The rationale is that if the sum of ranks belonging to the set of lower ranks is much larger than that in the set of higher ranks, the attribute is likely to be significant. For two level attributes, assuming that we group the ranks into {1,2,3,4} and {5,6,7,8}, the largest possible difference between the sums of ranks in the two groups is 16, i.e., 26 minus 10. Using the 67 percentile cutoff point, we identified cases in which the difference is equal to or greater than 12 as "likely-to-be-significant." For three level attributes, assuming that we group the ranks into {1,2}, {3,4,5,6}, and {7,8}, the largest possible difference between the sums of ranks in the most preferred and the least preferred rank sets is 12, i.e., 15 minus 3. Again, using the 67 percentile cutoff point, we identified the cases in which the difference is equal to or greater than 8 as "likely-to-be-significant." By comparing attributes actually found significant by various tests with those "likely-to-be-significant" identified directly from the data, we checked the face validity of the alternative tests.

According to <Table 8>, our *h*-test and Kohli's test show similar performance in terms of face validity. Among the 43 cases, significant attributes of the *h*-test were the same as the "likely-to-be-significant" attributes in 25 cases. If we include the cases that found one of the two "likely-to-be-significant" attributes to be significant (e.g., Case 22) or those that found the "likely-to-be-significant" attribute to be

significant along with another attribute (e.g., Case 5), the number increases to 35. These numbers change to 27 and 35, respectively, if we exclude attributes considered to have interactions rather than main effects. The equivalent numbers for Kohli's test are 27 and 34, respectively.

Although the two tests show similar performance in terms of face validity, their test results are not identical. In Cases 5, 24, 27 and 36, the *h*-test finds DISTANCE, MATE or UNDER to be significant while Kohli's test does not. Kohli's test is in accordance with the "likely-to-be-significant" attributes in these cases. Some of the attributes, for example DISTANCE in case 5 and MATE in case 36, are likely to have interaction effects with others. In Case 1 and 21, significant attributes of the *h*-test are in accordance with the "likely-to-be-significant" whereas those of Kohli's test are not. There are other cases, i.e., Cases 6, 8, 9, 16, and 20, where significant attributes are different between the two tests but become identical when we exclude potential interaction cases.

Unlike the *h* or Kohli's test, the *t*-test (MATE, UNDER, DISTANCE) and *F*-test (RENT) in regression finds too many significant variables. In 37 out of 43 cases, it finds all the "likely-to-be-significant" attributes to be significant. However, in 19 of the 37 cases, it also finds other attributes than the "likely-to-be-significant" to be significant.

Thus, regression analysis may be adequate if we would like to be generous in finding significant attributes. However, it may not be appropriate when we need to find a few critical attributes affecting the preference of a consumer. In benefit segmentation, we typically would like to identify a few critical attributes. Results of the Wald chi-square test for ordered logit are also shown in the table. It is clear that ordered logit analysis is not appropriate for individual level Conjoint analysis.

6.2 Comparison of the predictive power of models with significant attributes

An extensive empirical analysis was performed to compare the predictive power of models that incorporate only the significant attributes found by alternative tests. We relied on three different measures in order to compare the predictive power.

First, for each respondent and for each

testing method, we predicted the ranks of the eight estimation profiles with models that include only the significant attributes found in each test. In predicting ranks, we relied on all three estimated models. Then, we calculated the average correlations between the predicted ranks and the actual ranks over the three estimation models. <Table 9> shows the average correlations across respondents for each test method. As expected, regression analysis shows the highest correlation, having many significant attributes. The proposed *h*-test and Kohli's test show similar levels of correlation.

Second, for each respondent and for each testing method, we calculated errors after predicting ranks of the estimated data only with significant attributes. Then Mean Absolute Errors (MAE) and Mean Squared Errors (MSE) were calculated. <Table 9> shows the MAE scores. The pattern of MSE scores is very similar to that of MAE. Having more significant attributes, regression shows the best fit as expected. Again, the *h* and Kohli's test

<Table 9> Empirical Comparison with Other Test Methods for Ranking Data

	Correlation(ρ) with Significant Attributes	MAE for Main Profiles	MAE for Holdout Profiles
h-test (1)	0.887	1.104	0.969
h-test (2)	0.886	1.235	0.967
Kohli's test	0.878	1.271	1.009
Regression	0.913	0.785	0.683
Ordered Logit	0.809	1.930	1.450

* The average correlations between actual ranks and predicted ranks without significant attributes for each test.

show similar performance, although the h -test has a slightly lower MAE score.

Third, for each respondent and for each testing method, we calculated errors after predicting ranks of the hold out data using only the significant attributes. Then, for each method, we derived MAE and MSE. The MAE scores are reported in <Table 9>. Again the pattern is the same. Regression shows the best predictive power. The h test and Kohli's test show very similar performance in terms of the predictive power in general.

6.3 A brief comment on the analysis based on rating data

Among the 113 respondents, 53 respondents rated the fourteen profiles. For rating data, it is most appropriate to apply regression analysis to estimate part-worth utilities and test the significance of the attributes.

Nevertheless, we also estimated the parameters using MONANOVA and ordered logit. Next, we applied the t -test, h -test, Kohli's test and Wald test. For the h -test as well as Kohli's, we transformed the rating data into ranking data to find significant attributes. For the profiles having the same ratings, average ranks were assigned. Following the same procedure described for ranking data, we eliminated 11 cases that showed low reliability resulting in 42 responses for further analysis.

Here, we did not identify the "likely-to-

be-significant" attributes from the raw data. The attributes found to be significant by regression analysis are the most valid ones if we assume that the ratings are indeed interval scaled. The h -test and Kohli's test identify fewer attributes than regression to be significant. <Table 10> shows the significant attributes for each respondent for each test method.

In <Table 10>, by comparing significant attributes found in the h or Kohli's test with those found in the t -test, we see the face validity of the two tests are about the same. Although we do not present detailed results, we also confirmed that the predictive performance of the h and Kohli's test are similar. In the rating data, there are many profiles that have the same ratings. Because the h test and Kohli's test rely on ranking data, information could be lost in the process of transforming the rating data.

VII. Discussion

Formal testing methods to assess attribute significance are needed for nonmetric conjoint models. We proposed a new chi-square test to assess the significance of attributes. The test is simple, easy to use, can be applied at the individual level as well as at the aggregate level, and can be employed for the trade-off approach as well as for the full profile

<Table 10> Significant Attributes with Various Tests for Rating Data

Respondent	h-test	Kohli's test	Regression (t-test)	Ordered Logit (Wald χ^2 test)
1	RENT	NONE	NONE	MATE, RENT
2	MATE	MATE, DISTANCE	MATE	NONE
3	RENT	RENT	RENT	NONE
4	MATE	MATE	MATE,RENT	NONE
5	MATE	MATE	MATE,UNDER	NONE
6	RENT	RENT	MATE,RENT	NONE
7	MATE	MATE, DISTANCE	MATE, DISTANCE, RENT	NONE
8	MATE	MATE	MATE,UNDER	NONE
9	MATE	MATE	MATE	NONE
10	MATE	MATE	NONE	NONE
11	RENT	NONE	MATE,RENT	NONE
12	NONE	RENT	RENT	NONE
13	MATE	MATE	NONE	NONE
14	RENT	MATE	NONE	NONE
15	MATE	MATE, DISTANCE	MATE	NONE
16	RENT	MATE	MATE,RENT	NONE
17	NONE	DISTANCE	MATE, DISTANCE, RENT	NONE
18	RENT	RENT	MATE, RENT	NONE
19	NONE	NONE	MATE	MATE, RENT
20	MATE	MATE	MATE	MATE, RENT
21	MATE	MATE	MATE,RENT	NONE
22	MATE	MATE, DISTANCE	MATE	NONE
23	MATE	MATE	MATE	NONE
24	MATE, RENT	MATE, DISTANCE	MATE,DISTANCE, UNDER, RENT	NONE
25	MATE	MATE	MATE	NONE
26	RENT	RENT	MATE, UNDER, RENT	NONE
27	NONE	NONE	RENT	NONE
28	NONE	MATE	MATE, UNDER, RENT	NONE
29	NONE	NONE	NONE	RENT
30	MATE	MATE	MATE	NONE
31	NONE	DISTANCE	NONE	NONE
32	NONE	NONE	NONE	NONE
33	NONE	MATE	MATE	NONE
34	MATE	MATE, DISTANCE	MATE	NONE
35	MATE	MATE, DISTANCE	MATE	NONE
36	MATE	MATE, DISTANCE	MATE,RENT	NONE
37	RENT	NONE	RENT	MATE, RENT
38	RENT	NONE	RENT	NONE
39	RENT	NONE	NONE	MATE, RENT
40	NONE	RENT	NONE	NONE
41	MATE, RENT	MATE, DISTANCE	MATE,DISTANCE, RENT	NONE
42	NONE	MATE	MATE,RENT	NONE

* Significant at $\alpha = 0.10$

** MATE: Number of roommates. DISTANCE: Distance to class. UNDER: Undergraduate allowed

approach. Furthermore, by determining an appropriate number of sets of rankings, researchers can make the test more conservative or less conservative depending on the testing situation. We also illustrated how the test can be used and compared the face validity as well as the predictive performance of the method with other alternatives in an empirical application.

In general, the proposed method and Kohli's test show similar performance in terms of face validity and predictive power. Both methods appear to identify the real critical attributes to be significant unlike regression analysis. One advantage of the proposed method compared to Kohli's test is in its flexibility. Depending on the determination of number of subsets of ranks, the test can be made more conservative or less conservative. The flexibility could be also be considered a weakness of the proposed test compared to Kohli's test. Because Kohli's test relies on the sum of ranks, it may be easier to compute the test statistic. Determining an appropriate number of subsets may be an added burden for researchers.

A second advantage of the proposed test lies in its ability to identify potential interactions. Because it constructs contingency tables with attribute levels and rankings in the process of testing, it is easier to identify interactions of an attribute with other attributes. Once we identify an attribute to be significant, we should check whether the significance is due to

the main effect or the potential interaction effect.

Our numerical analysis using a few hypothetical cases reveals that sometimes a more generous α level may be needed to apply the proposed test or Kohli's test. When we had only 8 profiles for example, the proposed test identified only one significant case if we define 2 or 4 subsets of ranks and $\alpha = .10$. However, with a slightly expanded critical region, the test became very useful in identifying significant attributes for reasonable cases. This is also true of Kohli's test.

For many practical applications of the nonparametric Conjoint Analysis, importance weights are used to assess the importance of attributes. However, the importance weight is not based on any statistical theory. It is very sensitive to the range and number of attribute levels. The h -test proposed in this paper as well as Kohli's test are better alternatives when researchers need more theoretically grounded statistical tests.

〈논문 접수일: 2007. 03. 19〉

〈게재 확정일: 2007. 05. 10〉

References

- Cattin, P. and D. R. Wittink(1982), "Commercial Use of Conjoint Analysis: A Survey," *Journal of Marketing*, 46 (Summer),

- 44-53.
- Friedman, M.(1937), "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association*, 32 (December), 675-701.
- Green, P. E. and V. Srinivasan(1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5(September), 103-123.
- Green, P. E. and V. Srinivasan(1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *Journal of Marketing*, 54 (October), 3-19
- Johnson, R. M.(1974), "Trade-off Analysis of Consumer Values," *Journal of Marketing Research*, 11(May), 121-127.
- Kivetz, R., O. Netzer, and V. Srinivasan(2004), "Alternative Models for Capturing the Compromise Effect," *Journal of Marketing Research*, 41(August), 237-257.
- Kohli, R.(1988), "Assessing Attribute Significance in Conjoint Analysis: Nonparametric Tests and Empirical Validation," *Journal of Marketing Research*, 25(May), 123-133.
- Kruskal, W. H. and W. A. Wallis(1952), "Use of Ranks on One Criterion Variance Analysis," *Journal of the American Statistical Association*, 47(December), 583-621.
- Louviere, J. J., D. A. Hensher, and J. D. Swait (2000), *Stated Choice Methods: Analysis and Application*, New York: Cambridge University Press.
- Wittink, D. R. and P. Cattin(1989), "Commercial Use of Conjoint Analysis: An Update," *Journal of Marketing*, 53(July), 91-96.

〈Appendix〉

Mathematical Presentation of the New Method for the Tradeoff Approach

Assume that there are J attributes. We choose a focal attribute j ($j=1, \dots, J$) that has L levels ($l=1, \dots, L$) and another attribute j' that has M levels ($m=1, \dots, M$). There are $(J-1)$ design matrices that include attribute j . Denote the design matrices by $k=1, \dots, (J-1)$. Let $X_{l,m}^k$ be the cell where level l of the attribute j and level m of attribute j' appear in the design matrix k and $R(X_{l,m}^k)$ be the rank of the cell $X_{l,m}^k$.

We define disjoint subsets of ranks. Our experience suggests that it is appropriate to set the number of subsets equal to the number of levels of the focal attribute. Thus, we assume that ranks are grouped into the number of levels for the focal attribute.

$$S_q^k = \{(q-1)M + 1, (q-1)M + 2, \dots, (q-1)M + M\}$$

for $q=1, 2, \dots, L$

$$\text{Let: } f_{lq}^k = \sum_{m=1}^M I(R(X_{l,m}^k))$$

$$\text{where } I(R(X_{l,m}^k)) = \begin{cases} 1, & \text{if } R(X_{l,m}^k) \in S_q^k \\ 0, & \text{otherwise} \end{cases}$$

The f_{lq}^k represents the number of cells with level l of the focal attribute j that have ranks in the subset S_q^k in the design matrix k . Then,

$$\text{we define: } f_{lq} = \sum_{k=1}^{J-1} f_{lq}^k.$$

Finally, the f_{lq} 's can be arranged in the contingency table F :

$$F = \begin{Bmatrix} f_{11} & \dots & f_{1L} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & f_{lq} & \dots \\ \vdots & \ddots & \vdots & \dots \\ f_{L1} & \dots & f_{LL} \end{Bmatrix}$$

We apply the chi-square test using the contingency matrix F . The null hypothesis states that the attribute levels and the rank sets are independent.

$$H_0: P_{lq} = P_l P_q, \quad l=1, 2, \dots, L, \quad \text{and } q=1, 2, \dots, L$$

where P_{lq} is the probability that level l appears in the rank set q of the population, and P_l and P_q are marginal probabilities of level l and the rank set q appearing in the population.

The test statistic h is simply the chi-square computed as:

$$\sum_l \sum_q \frac{(f_{lq} - E[f_{lq}])^2}{E[f_{lq}]} \sim \chi^2((L-1)^2)$$

컨조인트 모형의 속성 유의성을 검증하기 위한 새로운 비모수통계 검증법

한 민 희*

Krishnamurthi, Lakshman**

강 현 모***

현 진 석****

박 상 준*****

현 용 진*****

국문요약

이 논문은 비모수 자료를 활용하는 컨조인트 모형에서 속성의 유의성을 평가하기 위한 새로운 카이제곱 검증법을 제안한다. 이 검증법의 가장 핵심적 아이디어는 수집한 서열을 몇 개씩 묶어 하위 집합들로 분류한 다음, 이 서열의 하위집합들과 속성별 수준들과 상호 통계적으로 독립적인지를 검증한다는 것이다. 이 검증은 프로파일들의 서열이 검증하는 속성의 수준에 대해 무작위로 분포되어 있다라는 명제를 귀무가설로 제시한다. 이 논문에서 새롭게 제시하는 검증방법은 매우 단순하고, 이해하기 쉬우며 사용하기도 쉽다. 이 방법은 총괄적 자료뿐 아니라 개인별 자료로도 검증이 가능하다. 또한 전체프로파일 법에서도 적용 가능하지만 더 나아가 트레이드오프 법에서도 적용이 가능하며, 현재로는 트레이드오프 법에 적용 가능한 유일한 통계적 검증방법이라 할 수 있다.

핵심개념: 비모수, 컨조인트, 카이제곱 검증법, 트레이드오프 법

* KAIST 경영대학원

** Kellogg Graduate School of Management

*** KAIST 경영대학원

**** DMS Lab(주)

***** 전북대학교

***** KAIST 경영대학원