

속성 병합에 의한 컨조인트 분석

임용빈[†] · 박가희 · 정종희

이화여자대학교 통계학과

Conjoint analysis by merging attributes

Yong B. Lim[†] · Gahee Park · Jong Hee Chung

Department of Statistics, Ewha Womans University

ABSTRACT

Purpose: A large number of attributes with mixed levels are often considered in the conjoint analysis. The respondents may have difficulty with scoring their preferences accurately because of many attribute items involved in each survey question. We research on the technique for reducing the number of attribute items.

Methods: In order to reduce the number of attribute items in a survey question, we make a new attribute by merging two original attributes. A 'No question' option is also included as a new level in a merged attribute.

Results: We propose BIB 6⁴ design in the case where we have four attributes with 2 levels and 3 levels, respectively and then analyze all the respondents survey data generated by the repeated simulation study in order to compare various model selection methods.

Conclusion: How to reduce the number of attribute items is proposed and how to design and analyze the survey data are illustrated.

Key Words: Conjoint analysis, Merging attributes, Variable selection methods and model selection criteria, BIB fractional factorial design

● Received 7 February 2016, 1st revised 17 March 2017, accepted 18 March 2017

[†] Corresponding Author(yblim@ewha.ac.kr)

© 2017, The Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and re-production in any medium, provided the original work is properly cited.

*This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A2002032)

1. 서 론

컨조인트 분석은 제품의 속성들과 해당 수준 조합으로 구성된 문항들을 만들어서 소비자들의 종합 선호도를 조사하여 분석하는 기법으로 주로 기업의 마케팅 분야에서 신제품의 콘셉트를 결정할 때 사용된다. 고전적인 컨조인트 분석은 응답자들 간에 동일한 프로파일, 즉 동일한 설문지를 사용한다. 응답자 한 명의 자료로 모델링을 하고 응답자마다 효과 크기를 계산한다. 본 연구에서는 임용빈 등(2015)에서 제안한 응답자들 간의 설문지 문항을 다르게 설계하는 방법을 사용한다. 핵심적인 아이디어는 설문지 문항들과 실험계획법의 실험점을 대응시키기 위해 응답자를 블록으로 간주하는 것이다. 즉 블록의 수가 응답자의 수가 되고 블록의 크기가 문항 수가 된다. 응답자들이 서로 다른 설문지에 응답하고, 응답자를 블록으로 간주하여 설문지들을 하나의 설계로 합쳐서 분석을 진행한다. 또한 선호도는 점수조사를 가정한다. 컨조인트 분석을 할 때 마케팅 담당자들은 정확한 응답과 속성들 간의 시너지 효과 또는 상충적 효과의 존재 여부까지 파악하는 것을 원한다. 즉 주효과와 이인자 교호작용효과를 모두 추정하는 것에 관심이 있다. 따라서 균형된 불완전 블록 완전요인설계(Balanced Incomplete Block Full Factorial Design)이거나 해상도가 V인 균형된 불완전 블록 일부요인설계(Balanced Incomplete Block Fractional Factorial Design)를 사용해서 설문지 문항설계를 한다. 조사하고자 하는 속성들의 수준수가 다른 경우 임용빈 등(2016)은 블록화 혼합수준 요인설계를 사용해 설문지 문항설계를 하고, 분석하는 방법을 제안하였다. 그러나, 블록화 혼합수준 요인설계를 사용할 때 고려하고자 하는 속성들이 많으면 한 문항 당 문항 되는 속성의 수가 많아지고 응답자들은 자신의 선호도를 정확하게 점수화하기 어려워질 수 있다. 즉 비표본 오차(non-sampling error)가 생길 수 있다.

본 연구에서는 컨조인트 분석에서 속성들의 수가 많은 블록화 혼합수준 요인설계를 사용할 때, 속성들의 병합을 통해 문항 당 문항 되는 속성의 수를 줄여 응답의 정확도를 높일 수 있는 방법에 대해 제안한다. 수준수가 서로 다른 속성 두 개와 ‘문지 없음’을 병합하여 새로운 속성으로 간주해서 설계를 생성하고 원래의 속성들의 수준으로 다시 전환하여 분석을 위한 설계를 생성한다. 이후의 절차는 기존의 블록화 혼합수준 요인설계의 절차와 동일하다. 제안한 방법을 사용할 경우 한 문항에서 문항 되는 속성의 수가 전체 속성의 수의 절반이하로 줄어들기 때문에 응답자에게 더욱 일관된 응답을 기대할 수 있다.

2절에서는 문항 설계에서 제안한 방법을 적용하는 과정을 소개하고, 3절에서는 유의한 속성을 선별하는 변수 및 모형 선택법에 대한 설명을 한다. 4절에서는 2수준과 3수준 속성이 각각 4개인 $2^4 \times 3^4$ 설계에 대하여 속성 병합 문항 설계 방법을 적용하여 설계를 생성하고, 속성들 간의 시너지 효과가 있는 모형을 가정한 후에 반복 시뮬레이션을 통해서 3절에서 소개한 변수 및 모형 선택법의 효율성을 비교한다.

2. 속성 병합 설계

수준 수는 다르지만 속성수가 같은 블록화 $a^f \times b^f$ 혼합설계인 경우, 임용빈 등(2016)은 블록화 a^f 와 블록화 b^f 설계의 교적설계로 문항설계를 제안하였다. 이때 모든 주효과와 이인자 교호작용효과를 추정하면서 응답자별 문항수를 줄이기 위해 수준수를 동일하게 만드는 방법이나, 일부 추정이 불가능한 이인자 교호작용효과를 허용하는 최소차수(minimum aberration)방법을 제안하였다. 이러한 방법들로 적절한 문항수를 제시할 수 있었으나, 한 문항을 구성하는 속성의 개수를 줄일 수는 없었다. 문항별 속성의 개수를 줄이면 응답자의 고려사항이 줄어들게 되어 응답자에게 더욱 일관된 응답을 기대할 수 있어서 설문조사에 대한 비표준 오차가 감소되는 효과를 기대할 수 있고, 설문조사의 신뢰성 제고에 영향을 미칠 것이라 기대된다. 따라서 본 연구에서는 속성 병합 설계를 제안하여 한 문항을 구성하는 속성의 개수를 줄일 수 있는 방안을 제시하고자 한다.

수준수 a 를 갖는 속성들의 수와 수준수 b 를 갖는 속성들의 수가 동일하게 f 인 문항 설계를 할 경우, a 수준 속성 1개와 b 수준 속성 1개 그리고 ‘묻지 않음’을 병합하여 $(a+b+1)$ 수준 속성 1개를 만들 수 있다. ‘묻지 않음’이란, 그 속성에 대하여 질문하지 않겠다는 뜻이다. 이러한 새로운 수준의 속성을 모두 f 개 만들 수 있으므로, $(a+b+1)^f$ 의 실험설계에 대하여 모든 주효과와 이인자 교호작용효과를 추정할 수 있으면서 블록크기를 최소화하는 균형된 불완전 블록 완전요인설계(Balanced Incomplete Block Full Factorial Design)이거나 해상도가 V 인 균형된 불완전 블록 일부요인설계(Balanced Incomplete Block Fractional Factorial Design)의 방법으로 문항 설계를 하는 것을 속성 병합 설계라 한다. 기존의 혼합설계에서는 한 문항 당 $2f$ 개의 속성 수준들이 결합되어 있었으나, 속성 병합 설계를 하면 최대 f 개에서 최소 0개로 한 문항 당 포함되는 속성들의 수를 줄이는 효과를 얻게 된다.

예를 들어 $a=2$, $b=3$ 이고 $f=4$ 인 경우, 즉 설문지 문항들이 2수준 속성 4개와 3수준 속성 4개의 수준값들로 구성된 경우에 속성 병합 설계를 구해보자. 2수준 속성과 3수준 속성을 병합하고 ‘묻지 않음’이란 새로운 수준값을 추가하면 $(2+3+1)^4$ 즉 6^4 설계를 고려해야 하는데 이때 6수준 속성의 첫 번째 수준은 ‘묻지 않음’, 두 번째와 세 번째는 2수준 값, 그리고 나머지는 3수준 값에 대한 것으로 병합한다. 그런데, 수준수 6이 소수(prime number)가 아니어서 블록화 6^4 설계를 직접적으로 생성할 수가 없다. 이 경우에 $6^4 = 2^4 \times 3^4$ 이기에 블록화 $2^4 \times 3^4$ 설계를 먼저 만든 다음 2수준과 3수준의 조합으로 6수준을 생성해서 블록화 6^4 설계를 만든다. 2수준 속성을 A,B,C,D로, 3수준 속성을 U,V,W,X로 표시하고, 알파벳 순서대로 2수준 속성과 3수준 속성이 각각 병합된 속성 4개를 AA, BB, CC, DD라 표시하자. SAS의 PROC FACTEX를 사용하여 블록화 6^4 설계를 생성하는 프로그램은 다음과 같다.

```
/*2수준 실험점 생성*/
proc factex;
factors A B C D/nlev=2;
blocks nblocks=4;
model r=max/minabs;
examine confounding aliasing ;
output out=d1 A nvals=(1 2)
           B nvals=(1 2)
           C nvals=(1 2)
           D nvals=(1 2);
run;
```

```
/*3수준 실험점 생성*/
proc factex;
factors U V W X/nlev=3;
blocks nblocks=9;
model estimate=(U|V|W|X @2);
examine confounding aliasing ;
output out=d2 designrep=d1 U nvals=(1 2 3)
           V nvals=(1 2 3)
           W nvals=(1 2 3)
           X nvals=(1 2 3);
run;
proc print data=d2; run;
```

```
/*6수준 실험점 생성*/
data six;
set d2;
AA=(A-1)*3+U;
BB=(B-1)*3+V;
CC=(C-1)*3+W;
DD=(D-1)*3+X;
block=(BLOCK-1)*9+BLOCK2
keep block AA BB CC DD;
run;
```

따라서, <Table 1>과 같이 블록화 6^4 설계의 블록개수 및 블록크기도 블록화 $2^4 \times 3^4$ 설계와 동일한 결과를 얻게 된다. 모든 이인자 교호작용효과가 추정 가능한 블록화 $2^4 \times 3^4$ 설계로부터 블록화 6^4 설계를 만들면 블록크기가 72로 한 사람당 문게 되는 설문지 문항수가 많아서 응답자로부터 각 문항에 대한 정확한 응답을 얻기 힘들기 때문에 추정할 수 없는 이인자 교호작용효과의 수를 최소로 하는 최소차수 설계를 사용해서 블록크기를 줄인다. 최소차수 방법을 적용한 블록화 6^4 설계는 블록크기가 36으로 줄어들게 된다.

블록화 6^4 설계에 대하여 분석을 하면, 병합된 속성 4개의 주효과와 이인자 교호작용효과들이 추정된다. 그러나, 분석자의 실제 관심은 2수준 4개 속성과 3수준 4개 속성 전체에 대한 주효과 및 이인자 교호작용효과일 것이다. 따라서, 분석을 할 때는 설계를 다시 2수준 속성 4개와 3수준 속성 4개로 전환하여 표현한다. 이렇게 바꾸는 방법은 <Equation (1)>과 같다. <Equation (1)>에 의하면 실제 2수준 속성이었던 A, B, C, D는 ‘묻지 않음’인 0이 포함되어 3수준 속성이 되고, 마찬가지로 실제 3수준 속성이었던 U, V, W, X도 4수준 속성이 된다. 이 경우에 원래 속성들의 모든 주효과와 병합된 속성들 간의 교호작용효과를 제외한 나머지 교호작용효과들을 추정할 수 있게 된다.

Table 1. Block Factorial Designs for $2^4 \times 3^4$ and 6^4
(N : number of runs, p : number of blocks, q : block size)

Block Design	N	p	q	design & block generator	Inestimable effects
$2^4 \times 3^4$	1,296	18	72	$b_1 = ABCD$ $b_2 = VWX, b_3 = UW^2X$	-
$2^4 \times 3^4$	1,296	36	36	$b_1 = CD, b_2 = ABD$ $b_3 = VWX, b_4 = UW^2X$	CD
6^4	1,296	18	72	$b_1 = ABCD$ $b_2 = VWX, b_3 = UW^2X$	AU,BV,CW,DX
6^4	1,296	36	36	$b_1 = CD, b_2 = ABD$ $b_3 = VWX, b_4 = UW^2X$	WX,AU,BV,CW,DX

$$\begin{aligned}
 A &= \begin{cases} 0, AA = 1, AA \geq 4 \\ 1, AA = 2 \\ 2, AA = 3 \end{cases} & B &= \begin{cases} 0, BB = 1, BB \geq 4 \\ 1, BB = 2 \\ 2, BB = 3 \end{cases} \\
 C &= \begin{cases} 0, CC = 1, CC \geq 4 \\ 1, CC = 2 \\ 2, CC = 3 \end{cases} & D &= \begin{cases} 0, DD = 1, DD \geq 4 \\ 1, DD = 2 \\ 2, DD = 3 \end{cases} \\
 U &= \begin{cases} 0, AA \leq 3 \\ 1, AA = 4 \\ 2, AA = 5 \\ 3, AA = 6 \end{cases} & V &= \begin{cases} 0, BB \leq 3 \\ 1, BB = 4 \\ 2, BB = 5 \\ 3, BB = 6 \end{cases} & W &= \begin{cases} 0, CC \leq 3 \\ 1, CC = 4 \\ 2, CC = 5 \\ 3, CC = 6 \end{cases} & X &= \begin{cases} 0, DD \leq 3 \\ 1, DD = 4 \\ 2, DD = 5 \\ 3, DD = 6 \end{cases}
 \end{aligned} \tag{1}$$

3. 적절한 모형 찾기

2절에서 속성 병합 설계를 생성하고 반응변수인 선호도 점수에 대한 분석을 위해 각 문항을 원래 속성들의 수준값으로 표현하였다. 적절한 모형 찾기를 위하여 잘 알려진 축차적인 변수 선택법인 전진적 선택법(forward selection), 후진적 제거법(backward elimination)과 단계적 회귀(stepwise regression) 방법을 이용하고, 변수를 선택하는 기준은 유의수준, 본페로니 보정된 유의수준, AIC, AICc와 BIC 를 적용한다. 각 모형 선택 기준에 대한 설명은 다음과 같다.

m개의 검정을 동시에 진행할 때(multiple comparison), 동시 검정에 대한 제1종 오류의 확률을 유의수준 α 로 유지하기 위해 각 검정에 대한 유의수준 값으로 α/m 를 사용하는 것을 본페로니 보정(Bonferroni correction) 방법이라 한다.(Bland and Altman, 1995) 변수 선택을 할 때에도 m개의 변수 각각의 계수가 유의한지에 대한 검정을 동시에 진행하기 위해 전진적 선택법의 각 단계에서 변수를 추가할 때 본페로니 보정에 의한 수정된 유의수준 값인 α/m 을 사용해서 변수를 추가할 지를 결정한다. 변수를 추가할 때 사용되는 유의수준 값이 단계마다 작아진다는 것을 제외하고 전진적 선택법과 동일하다. 모형에 포함된 변수가 많을수록 더 작은 유의수준 값으로 검정하기 때문에 해당 방법은 보수적으로 변수를 선택하게 된다.

AIC(Akaike's Information Criterion)는 적합한 모형을 찾는 지표로 우도함수로 모형의 적합도를 평가하고, 추정된 모수의 개수로 모형의 복잡성에 대한 패널티를 주어서 계산한다. 계산식은 <Equation (2)>와 같고 값이 작을수록 좋은 모형이라고 판단한다. <Equation (2)>에서 <Equation (4)>까지 사용된 n은 자료 수이고, p는 추정된 모수의 개수, \hat{L} 는 우도비 함수의 최대값을 나타낸다.

$$AIC = -2 \cdot \ln \hat{L} + 2 \cdot p \quad (2)$$

실험크기가 작은 설계에서 AIC는 너무 많은 항을 선택하는 경향이 있다. 때문에 크기가 작은 설계에서 추가적인 항에 대한 패널티를 더 크게 줌으로써 이를 극복하고자 한 방법이 AICc(corrected Akaike's Information Criterion) 방법이고 <Equation (3)>과 같다.

$$AICc = AIC + \frac{2p(p+1)}{n-p-1} \quad (3)$$

BIC(Bayesian Information Criterion) 기준은

$$BIC = -2 \cdot \ln(\hat{L}) + \ln(n) \cdot p \quad (4)$$

와 같이 정의되는데, AIC에서는 모형의 복잡성에 대한 패널티를 2p만큼 주는데 반해 BIC의 경우 $\ln(n) \cdot p$ 만큼의 패널티를 줘서 AIC보다 간결한 모형을 선택하게 된다.

4. 시뮬레이션을 통한 모형 선택법 비교

2수준과 3수준 속성이 각각 4개인 $2^4 \times 3^4$ 문항 설계를 위하여 2절에서 소개한 속성 병합 설계와 최소차수 설계를 이용하여 블록의 크기가 36인 블록화 6^4 설계를 생성하고, 각 문항을 <Equation (1)>에 따라서 해당 속성에 관해 묻지 않음을 0 수준 값으로 포함하여 원래 속성들의 수준값으로 표현한다. 각각의 문항에 대한 응답자들의 선호도 점수를 시뮬레이션 모형에 의해 생성하기 위해 주효과 A, C, V, W와 이인자 교호작용효과 AV, AW를 이용한 효용함수를 가정한다. 주어진 문항에 대한 선호도 점수는 <Equation (5)>와 <Table 2>에 주어진 가정된 효과의 크기에 따라서 생성된 후에,

$$y_{ijkl} = 30 + a_i + c_j + v_k + w_l + (av)_{ik} + (aw)_{il} + \varepsilon_{ijkl} \quad i, j = 0, 1, 2, \quad k, l = 0, 1, 2, 3 \quad (5)$$

$$\text{여기서 } \varepsilon_{ijkl} \sim N(0, \sigma^2)$$

1점에서 9점까지 최종 선호도 점수는 <Equation (5)>로부터 생성된 모든 y값의 분포에 따라 등간격으로 나눈 후 부여한다. 여기서 0 수준에서의 각 효과의 크기는 0이다. 또한 효용함수의 오차항은 평균이 0이고 표준편차가 σ 인 정규분포를 따르고 σ 가 각각 1, 1.5, 2인 세 가지 경우로 가정한다. 각 오차항의 가정에서 1,000번의 시뮬레이션 자료를 생성시켜 분석을 반복한다. 효율성의 지표인 power는 1,000번 중 오로지 가정된 효과만 유의한 효과들로 찾은 비율로 정의한다.

Table 2. Effect size assumption for blocked $2^4 \times 3^4$ design

level	a_i	c_j	v_k	w_l	$(av)_{1k}$	$(av)_{2k}$	$(aw)_{1l}$	$(aw)_{2l}$
1	-1	-1.5	-1	-1	-1.5	1.5	-1.5	1.5
2	1	1.5	0	0	0	0	0	0
3			1	1	1.5	-1.5	1.5	-1.5

시뮬레이션은 1,000번을 반복하였고 가정된 유의한 속성들만 잘 찾는 비율인 power의 결과 값은 <Table 3>과 같다. 시뮬레이션 결과를 살펴보면 power에서 가장 좋은 성능을 보이는 방법은 BIC를 기준으로 한 변수선택법과 본페로니 보정된 유의 수준값을 사용한 전진적 선택법이다. BIC를 기준으로 한 변수선택법의 경우 오차항의 표준편차 크기 가정에 관계없이 전진적 선택법, 후진적 제거법과 단계적 회귀법에서 모두 높은 power를 보인다. 또한 유의 수준을 0.01로 하는 본페로니 보정된 유의 수준값을 사용한 전진적 선택법 역시 높은 power를 보인다. BIC를 기준으로 한 변수선택법과 본페로니 보정된 유의 수준값을 사용한 전진적 선택법은 모두 보수적으로 변수를 선택하는 방법이라고 할 수 있는데, 두 변수선택법에서 모두 power가 매우 높게 나타났고, 이를 통해 적절한 모형찾기를 하였을 때 제안한 방법의 효율성이 높게 나타나는 것을 알 수 있다.

Table 3. Proposed blocked $2^4 \times 3^4$ design power

		$6^4_{(36)36}$		
		$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
AIC	Forward	0.271	0.269	0.282
	Backward	0.097	0.081	0.088
	Stepwise	0.271	0.269	0.282
AICc	Forward	0.396	0.404	0.413
	Backward	0.192	0.198	0.178
	Stepwise	0.396	0.404	0.413
BIC	Forward	0.999	0.998	0.997
	Backward	0.999	0.998	0.997
	Stepwise	0.999	0.998	0.997
Forward P-value	$\alpha = 0.01$	0.783	0.779	0.787
	0.05	0.284	0.271	0.269
	0.1	0.083	0.080	0.077
Backward P-value	0.01	0.778	0.775	0.775
	0.05	0.264	0.251	0.243
	0.1	0.068	0.065	0.066
Stepwise P-value	0.01/0.01	0.783	0.779	0.787
	0.05/0.05	0.284	0.271	0.269
	0.01/0.05	0.783	0.779	0.787
Forward Bonferroni	0.01	0.982	0.978	0.980
	0.05	0.914	0.901	0.908

5. 요약

본 연구는 컨조인트 분석에서 고려하고자 하는 속성의 수가 많은 경우 속성들의 병합을 통해 문항 당 묻는 속성의 수를 줄이는 방법을 제안한다. 속성 병합에 의한 문항 설계의 경우에 각 문항에 대한 응답자의 더욱 일관된 응답을 기대할 수 있어서 응답의 정확도를 높일 수 있다는 점에서 제안된 방법의 의의를 찾을 수 있다. 사례분석으로 2수준과 3수준 속성이 각각 4개인 $2^4 \times 3^4$ 문항 설계를 위하여 제안된 속성 병합 설계를 이용하여 블록의 크기가 36인 블록화 6^4 설계를 생성하고, 각각의 문항에 대한 응답자들의 선호도 점수를 시뮬레이션 모형에 의해 생성하기 위해 주효과 A, C, V, W와 이인자 교호작용효과 AV, AW를 이용한 효용함수를 가정한다. 축차적인 변수 선택법과 모형 선택 기준에 따라서 1000회 반복된 시뮬레이션 모형에 의해 생성된 문항 자료를 분석하여 각 방법의 power를 비교한 결과, 보수적으로 변수를 선택하는 방법인 BIC를 기준으로 한 변수선택법과 유의수준을 0.01로 하는 본페로니 보정된 유의 수준값을 사용한 전진적 선택법이 적절한 모형 선택 방법으로 추천된다.

속성 병합 설계를 사용하는 경우에는 병합된 속성들 간의 교호작용효과를 추정할 수 없게 된다는 단점이 있다. 특정한 속성들 간의 교호작용효과가 존재할 가능성이 높다고 알려진 경우에는 실무 담당자들이 속성 병합 설계시에 병합되는 속성들의 배치를 조절해 잠재적인 속성들의 이인자 교호작용효과들은 추정가능하게 함으로써 이 문제는 해결 할 수 있다.

REFERENCES

- Bland, J. M., and Altman, D. G. 1995. "Multiple significance tests: the Bonferroni method." *Bmj* 310(6973):170.
- Lim, Yong B., and Chung, Jong Hee. 2016. "Conjoint analysis with mixed levels of attributes." *Journal of the Korean society for Quality Management* 44(4):799-811.
- Lim, Yong B., Chung, Jong Hee, and Kim, Joo H. 2015. "Practical designs, analysis and concepts optimization in conjoint Analysis." *Korean Journal of Applied Statistics* 28(5):951-963.

