

명목척도를 갖는 변수의 축약방법에 관한 연구
(Reduction of Nominal Variables Using Factor Analysis Model)

홍순욱
영동대학교 건설공학부

조근태
Katz Graduate School of Business
University of Pittsburgh

권철신
성균관대학교 산업공학과

Abstract

In this article, a reduction method for nominal variables is presented and its use illustrated. Factor analysis model (FAM) generally enables us to reduce variables having interval or ratio scale based on their correlation coefficients. We developed an extensive method that makes FAM applicative to the case of nominal variables which does not give correlation coefficients, but only the degree of association. Cramer's V coefficient is a well-established measure that provides the strength of association for nominal variables with a range of [0,1]. When Cramer's V coefficient can logically substitute for correlation coefficient, FAM would be extensively used for reduction of nominal variables.

1. 서론

일반적으로 측정변수의 수준이 높을수록 정교한 통계적 모형을 분석에 적용할 수 있으나, 측정비용이 높은 단점이 있다. 명목척도는 측정이 용이한 반면, 분석에 제약이 많아 조사설계에 적지 않은 애로를 초래하는 실정이다. 본 연구에서는 요인분석 모형을 이용하여, 명목척도의 수준으로 측정된 변수들을 축약하기 위한 새로운 방법을 제안하고 있다. 즉, 요인분석은 변수간의 상관관계에 근거하여 수행된다는 점에 착안하여 Pearson의 상관계수와 같은 형태로서 명목척도를 갖는 변수간의 연관성(association)을 나타내는 Cramer's V계수를 요인분석의 입력 데이터로 채용하여 변수의 축약을 시도한다. 이와 같은 방법을 응용한다면 명목척도로 측정된 변수에도 등간척도 이상의 값을 갖는 변수와 마찬가지로 요인분석 등을 비롯한 상관관계의 개념에 근거하는 모수통계적 분석 방법을 앞으로 제한적

이나마 적용할 수 있을 것이다. 마지막으로 이것이 어떻게 적용되는가를 보이기 위하여 R&D 프로젝트에 관련된 변수들을 대상으로 한 간단한 분석사례를 제시한다.

2. 요인분석 모형

기본적인 요인분석 모형은 다음과 같다[2]. 즉, p개의 변수에 대한 확률벡터 \mathbf{X} 의 평균이 $\boldsymbol{\mu}$ 이고, 공분산행렬이 $\boldsymbol{\Sigma}$ 라고 할 때, \mathbf{X} 는 F_1, F_2, \dots, F_m 의 관찰불가능한 확률변수인 m개의 공통요인과 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ 의 p개의 개별요인에 대하여 선형으로 종속되어 있다고 본다. 이를 행렬의 형태로 표현하면 다음과 같다.

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L} \mathbf{F} + \boldsymbol{\varepsilon}$$

(p×1) (p×m) (m×1) (p×1)

여기서 행렬 \mathbf{L} 은 요인부하량 행렬이다. 만약 $\boldsymbol{\Sigma}$ 가 대각선행렬이 아니고 비대각원소의 크기가 크다면, 이 경우에 요인분석을 의미있게 적용할 수 있으며, 요인부하량과 개별분산의 대표적인 추정방법으로 주성분법과 최우추정법이 있다. 어느 것이든 관계 없지만, 주성분법을 채택한 요인분석 모형에서 본 연구와 관련된 주요특징을 보면 다음과 같다.

- (1) 변수수준: 등간 또는 비율척도
- (2) 입력자료: pearson의 상관계수행렬 또는 공분산행렬
- (3) 자료의 적합도: Kaiser-Myler-Olkin (KMO) measure, Bartlett test of sphericity
- (4) 회전방법: 요인간 독립성을 유지하기 위한 직각회전
- (5) 요인추출기준: 고유값 1.0 이상

여기서 요구되는 입력자료는 등간척도 이상의 변수들간에 존재하는 pearson상관계수행렬이다. 그러나 명목척도를 갖는 변수에서는 pearson상관계수가 존재하지 않으므로, 이 문제의 해결을 위하여 명목척도 변수간의 연관성을 나타내는 지표 가운데 요인분석에의 적용이 가능하면서, 동시에 pearson상관계수의 형태와 가장 근접한 지표를 찾아야 할 것이다.

3. Cramer's V 계수

흔히, 명목척도를 갖는 변수의 상관성을 파악하기 위하여 사용되는 방법은 χ^2 -test이다. 그러나 χ^2 통계량은 $[0, \infty]$ 의 값을 가질 뿐 아니라 변수간의 관련정도(strength of association)을 알 수 없기 때문에 χ^2 통계량을 기초로 하여 다음과 같이 개량된 척도들이 개발되어 왔다[5].

(1) 먼저, $[\chi^2/N]^{1/2}$ 로 정의되는 ϕ 계수로서, 표본크기 N으로 χ^2 통계량을 조정하고자 하는 것이지만, $\chi^2 > N$ 의 가능성이 존재하므로 ϕ 가 반드시 $[0, 1]$ 사이에 오지 않는다.

(2) $[0, 1]$ 사이의 값을 얻기 위하여 Pearson은 다음과 같은 상황계수(Coefficient of Contingency) $C = [\chi^2/(\chi^2 + N)]^{1/2}$ 를 제안하였다.

이 상황계수의 값은 $[0, 1]$ 사이에 존재하지만, 상한값(upper limit)이 1이 되지 않을 수 있다. 상황계수의 최대값은 교차분석표의 행과 열의 數에 영향을 받는다.

(3) 이 점을 극복하기 위하여 Cramer는 다음과 같은 Cramer's V 계수를 제시하였다. $V = [\chi^2/\{N(k-1)\}]^{1/2}$, k = (행의 개수와 열의 개수 중 최소치)

Cramer's V 계수를 사용하면 어떠한 차원(Dimension)을 갖는 교차표에서도 $[0, 1]$ 의 범위에서 명목척도로 측정된 두 범주형 변수의 상관성을 표현할 수 있다.

4. 적용절차

Cramer's V 계수와 요인분석 모형을 사용하여 명목척도 값을 갖는 변수를 축약하기 위한 절차는 다음과 같다.

- (1) 축약대상 범주형 변수의 결정
- (2) 교차분석 실시를 통한 Cramer's V 계수행렬의 산정
- (3) KMO measure와 Bartlett test를 이용한 자료의 적합도 검정
- (4) 요인분석의 실시
- (5) 요인추출 및 naming
- (6) 사후분석 및 조정

다음 장에서는 이와 같은 절차가 간단하게 적용된 例가 제시된다.

<표1> 설정된 변수

변수	description	scale	변수	description	scale
변수1	프로젝트 형태	전략과제, 수명과제, 위탁과제, 자체과제	변수5	기술발전 단계	도입기, 성장기, 성숙기, 쇠퇴기
변수2	프로젝트 산출	제품개발과제, 기술개발과제	변수6	팀장의 리더쉽	개방형, 권위형, 겸비형, 모호형
변수3	프로젝트 수준	첨단제품/첨단기술과제, 개량제품/개량기술과제	변수7	연구소 운영풍토	관리중심형, 자율중심형, 양면조화형, 모호형
변수4	프로젝트 단계	응용연구과제, 개발연구과제, 상품화연구과제	변수8	프로젝트 수행형식	task force형, project형, matrix형

5. 적용예

가. 변수의 설정

본 연구에서 제안된 명목척도를 갖는 변수의 측약 방법의 적용예를 보이기 위하여 사용된 변수 <표1> 는 R&D 프로젝트 관련 변수를 다룬 홍[4]의 연구에서 수집, 정리된 데이터를 근간으로 하고 있다. 자료의 케이스는 총329개로 우리 나라 전기·전자산업에 속한 기업의 18개 연구소를 대상으로 수집된 것이다. 통계분석에는 SPSS/PC+를 사용한다.

나. 적용의 결과

상기의 8개 변수에 대한 Cramer's V 계수행렬은 SPSS/PC+의 CROSSTABS Subroutine에서 제공되는 모두 28개($8C_2$)의 교차분석표에 나타난 데이터를 정리하면 얻을 수 있다. 이 행렬이 명목척도를 갖는 변수간의 상관성을 나타내는 데이터로서 요인분석에 적합한 것인가를 판정하기 위

한 적합성 검정을 실시한 결과, KMO 측정치는 0.66603, Bartlett 검정통계량 및 p-value는 각각 222.9833와 0.0000 으로 나타나, 주성분분석의 입력자료로서 적합하다고 판정된다. 이를 요인분석 모형에 입력하여 얻어진 요인적재행렬은 <표2>와 같다.

「요인1」의 경우, 프로젝트 단계는 응용연구, 개발연구, 상품화과제의 범주를 갖고 있어 개발과제의 기본적인 성격을 나타내고 있으며, 프로젝트의 산출은 목표하는 산출물의 형태 즉, 최종적인 개발성과가 제품인가 기술인가를 구분함으로써 역시 프로젝트의 성격을 알 수 있고, 프로젝트의 수행형식은 개발과제가 어떠한 조직의 형태로 추진되는지를 말해 주기 때문에 이들 조건변수로 구성된 요인은 프로젝트의 개략적인 성격과 특징에 대한 정보를 준다. 따라서 첫 번째 요인을 「프로젝트 특성」으로 파악할 수 있다. 「요인2」는 개발대상의 기술수준을 첨단성을 구분해 주며, 요구되는 기술이 기술수명주기 상의 어디

<표2> Cramer's V 계수를 사용한 요인적재행렬

요 인	요인 1	요인 2	요인 3
조건변수			
프로젝트 단계	.77400	.19926	.00562
프로젝트 산출	.76552	.22966	-.04202
프로젝트 수행형식	.63946	-.07830	.22507
프로젝트 수준	.04969	.79335	.03898
기술발전 단계	.19112	.73006	.01692
프로젝트 형태	.06946	.36705	.31308
연구소 운영풍토	.10702	-.05805	.76446
팀장의 리더쉽	.00602	.16485	.71480

에 위치하는지를 평가하게 해 준다. 또한 프로젝트의 형성원천을 top-down(전략과제 또는 수명과제), bottom-up(자체과제), contract(위탁과제) 등으로 구분해 준다. 일반적으로 개발활동이 첨단제품/기술을 대상으로 할수록, 기술의 태동기 쪽으로 위치할수록, 최고경영층이 전략적 목표를 갖고 추진한 프로젝트일수록 개발의 난이도가 높다고 판단되므로, 두 번째 요인은 「개발 난이도」로 명명할 수 있다. 「요인3」은 크게 두 가지 요소로 구성되어 있다고 하겠다. 첫째, 연구소 조직내부의 분위기, 즉 관리 중심으로 운영되는 풍토인가 아니면 자율 중심으로 이루어지는 것인가를 보는 것이고, 둘째, 프로젝트 리더의 성향이 개방적인지 권위적인지를 구분하여 준다. 이들 요소는 모두 연구의 동기, 활력 그리고 분위기와 밀접하게 관련되어 있으므로 「연구 환경」이라는 조건변수로 간주하여도 가능할 것이다.

이렇게 본다면 8개의 초기 조건변수는 Cramer's V계수와 요인분석을 이용함으로써 프로젝트의 특성, 개발 난이도, 연구 환경이라는 세 가지의 새로운 변수로 축약되었으며, 연구의 목적에 따라서 이들을 다양한 형태로 추가분석에 사용하는 것도 가능할 것이다.

6. 결론

본 연구에서는 요인분석 모형과 Cramer's V 계수를 이용하여 상호연관성을 갖는 명목척도 변수의 축약을 가능하게 해 주는 방법을 제시하였고, 우리나라의 기업연구소에서 수행되었던 연구개발 프로젝트에 관련된 연구변수의 일부를 이용하여 본고에서 제시된 방법이 구체적으로 어떻게 적용될 수 있는지 보였다.

이러한 방법의 개발로 범주형 변수의 취급 및 조작의 범위가 넓어지게 되었고, 이로 인하여 패턴이나 유형에 관심을 두는 탐색적 연구에 있어서 더욱 강화된 실증적 기반을 가질 수 있는 가능성을 열었다는 점에서 본 연구의 의의가 있다 하겠다.

앞으로, 본 방법을 사용하는 응용연구뿐 아니라, 이에 대한 이론적 고찰의 진전이 기대된다.

참고문헌

- [1] 권철신, 기술경영구조분석대전, 삼성전자, 1991
- [2] 김규형, 다변량분석의 이론과 실제, 중앙대학교 출판부, 1993
- [3] 채서일, 김범중, SPSS/PC+를 이용한 통계분석, 범문사, 1990
- [4] 홍순욱, 기술커뮤니케이션의 구조해명에 관한 연구, 성균관대학교 대학원 박사학위논문, 1992
- [5] Goodman, L.A., Kruskal, W.H., Measures of association for Cross-Classification, *Journal of the American Statistical Association*, Vol. 49, No.268, pp.732-764, 1954.
- [6] Hair, J.F., et al., *Multivariate Data Analysis with Readings*, Petroleum Publishing Company, 1979
- [7] Norusis, M.J., *SPSS/PC+ for the IBM PC/XT/AT*, SPSS inc., 1986