

「北韓의 行態 豫測을 위한 要因分析 (factor analysis)의 應用」

黃 東 準*

I. 序 論

現代社會는 情報의 社會이고 情報에 의한 政策의 싸움이 끊임없이 일어나고 있다. 한企業이 자기회사의 商品을 잘 팔기 위하여 競爭 會社의 販賣戰略의 탐색은 물론 消費者의 行態까지 면밀히 分析하여 새로운 商品을 開發하고 公害防止을 위한 政府의 강경한 방침에 따라 無公害 商品의 高안에 많은 投資를 하고 있는 것이 先進國의 實情이다. 다시말하면 未來의 狀況變化와 새로운 情報에 따라 最高經營者는 매일 매일 어려운 企業環境속에서 經營政策을 결정하고 있는 것이다.

하물며 한 國家의 發展과 生存을 左右하는 國家의 정책은 반드시 얻을 수 있는 모든 情報資料下에서 最適의 정책을 세워야 하고, 계속적인 評價와 修正 그리고 補完이 이루어져야 한다.

美國이 철군을 如作하고 美國과 中共의 國交가 정상화를 향하여 발돋움하고 있고 最近엔 美國의 三者會談 제안등 南北韓問題는 豫測을 불허하게 변하고 있다. 더구나 北韓은 200해리 경제수역은 물론 軍事수역까지 선언하는 등 우리 주변의 국제정세 역시 시시각각으로 변하고 있다.

무엇보다도 이러한 政治現象 및 軍事現象의 豫測에는 數值化할수 없는 定性的(qualitative) 要因들이 많기 때문에 豫測의 적절한 技法의 선택과 應用이 어려운 것이 사실이다. 시물레이 션모델이 豫測에 사용될 수 있겠고, 類推法, 델파이法(Delphi method), 時系列分析등의 方法

이 있겠으나, 여기에서는 國防問題 즉 南北韓關係 豫測에 효과적으로 적용될 수 있는 要因分析(factor analysis)에 대하여 살펴 보기로 한다.

II. 要因分析(factor analysis)豫測

要因分析은 南北韓關係의 未來狀況豫測에 應用될 수 있는 아주 좋은 計量分析法이다. 먼저 要因分析의 理論에대하여 살펴본 다음에 要因分析에 의하여 南北韓關係의 豫測에 적용할 문제를 알아 보기로 한다.

要因分析은 原則的으로 어떤 시스템을 構成하고 있는 수 많은 變數들을 몇개의 意味있고 해석할 수 있는 要因으로 減少시키는 方法이다. 要因分析은 다음과 같은 이유때문에 分析家나 未來豫測 專門家에게 훌륭한 分析 道具로서 잘 사용된다¹⁾

(1) 要因分析은 일련의 관측된 자료로부터 여러 變數間의 關係를 결정하여 주는 潛在要因(latent factor)을 규명하여 준다.

예를 들면 北韓이 南韓에 대하여 도발 事件을 지난 수십년동안 저질러 왔는데 이러한 도발사건들을 잔인성, 납치여부등 여러가지 변수로 구분하여 볼 수 있으나 이러한 사건들에 근본적으로 潛在해 있는 南韓의 政治 및 社會現象, 北韓의 國內與件, 그리고 주변 4強國의 국제관계등에 관한 어떤 特性(salient characteristics)들을 要因分析에 의하여 조사하여 불

(1) Wells, W.D., J.N. Sheth, "Factor Analysis in Marketing Research," in Handbook of Market Research, ed., Robert Ferber, McGraw-Hill Book Co., 1971

*陸軍士官學校

要性이 있다. 다시말하면 過去 資料를 要因分析함으로써 北韓의 도발형태를 豫測할 수 있고 分析된 南北韓 및 國際政治與件에 따른 特性에 따라 도발 強度도 알 수 있는 것이다.

(2) 要因分析은 관측된 資料로부터 쉽게 알 수 없었던 變數들간의 關係를 알게 하여 준다. 예를 들면 北韓의 도발은 분명히 南韓의 社會의 여건에 따라 그리고 美國의 태도에 따라갈 이 關係가 있는 줄은 모두 아는 事實이다. 그러나 그러한 여건(變數)들이 어떻게 關係되는 지를 알기는 쉽지가 않다. 要因分析은 이러한 때 분명히 存在하는 變數이지만 理解하기 어려운 諸關係를 알기 쉽게 分析케 한다.

(3) 要因分析은 어떤 變數들을 구분화할 필요가 있을때 사용된다. 要因分析으로 한 집단을 構成하고 여러 個人들을 共通된 要因(factor)으로 分類할 수도 있고 뒤에 설명하겠 지만 要因變數(factor score)에 따라 個人 또는 對象들을 구분화 할 수가 있다.

1. 要因分析이란 ?

앞에서도 說明하였으나 要因分析은 實在 관측된 전체 變數들간의 相互關係를 研究 分析하는데 사용되는 多變數 統計分析技法(multivariate statistical analysis technique)이다. 回歸分析(regression analysis)에서는 한 변수는 尺度變數(종속변수)로써 다른 모든 變數들은 推定(獨立)變數로써 취급하여 分析하나 要因分析에서는 모든변수들을 똑같이 취급하여 분석한다.

그러나 要因分析에서는 각 관찰된 변수들을 潛在的인 가상要因들의 종속함수로 고려한다. 反對로 말하면 各 要因들은 관찰된 변수들의 종속함수로 볼 수 있는 것이다.

要因分析의 主目的은 要因(factor)이라고 부르는 새로운 變數項目으로 여러 변수들을 分解하는 것이다. 要因은 관측된 변수들의 線型 關係이며 다음과 같이 쓸 수 있다.

$$F = a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n$$

여기에서 말하는 要因은 多重回歸分析(multiple regression analysis)의 종속변수와 같 다고 볼 수 있으며 各 變數 $x_j(j=1, 2, \dots, n)$

의 係數 $a_j(j=1, 2, \dots, n)$ 를 구하는 것도 多重 回歸分析에서 사용하는 最小自乘法과 똑같은 原理이다. 그러나 다른 點은 관측된 모든 變수들이 한 要因이 아닌 여러개의 要因들로 구분된다는 것이다. 즉 다음과 같은 關係가 要因分析에서는 나타난다.

$$\begin{aligned} F_1 &= a_{11}x_1 + a_{21}x_2 + a_{31}x_3 \\ F_2 &= a_{42}x_4 + a_{52}x_5 \\ F_3 &= a_{63}x_6 + a_{73}x_7 \end{aligned}$$

여기에서는 총 7개의 변수들이 3개의 要因으로 구분화된 것이다. 즉 처음 세변수 x_1, x_2, x_3 는 要因 F_1 에 다음의 x_4, x_5 변수는 要因 F_2 에, 그리고 마지막 변수 x_6, x_7 은 要因 F_3 에 구분된 것이다.

一般的으로 要因分析에서는 各變數(對象)에 要因點數(factor score)가 割當된다. 즉

$$F_i = a_{1i}x_1 + a_{2i}x_2 + a_{3i}x_3 + \dots + a_{ni}x_n$$

로 표시할 수 있다. 要因點數가 各 變數에서 관측된 點數와 相關關係가 있을때에 이러한 상관도를 “ 要因負荷(factor loading)”라고 한다. 만약 n 개 變數와 r 개 要因이 있을 경우에는 總($n \times r$)개의 要因負荷가 있으며 이것들은 “ 要因 負荷行列(factor loading matrix)”이라고 하는 行列內에 要約된다. 즉, 要因負荷行列은 관측된 變數와 要因들간의 相關도를 나타내는 行列인 것이다.

表 1은 要因負荷行列을 例를 들어 나타낸 것이다. 여기에서 各 要因의 負荷를 제공하여 합한 것을 自乘合(sum of square) 또는 그 要因의 eigenvalue라고 한다. 各 eigenvalue는 지금 고려하고 있는 6개 변수에 의한 總分散(total variance)에 대한 比率를 말한다. 變數들이 標準化(standardized)되면 各 변수들의 分散은 1이다. 따라서 表 1의 例에서 총분산은 變數의 수와 같은 6이므로, 要因 A는 전체 分散의 $2.89/6=48\%$ 를 그리고 要因 B는 $1.01/6=17\%$ 를 흡수하는 것을 의미한다. 따라서 要因 A와 B는 總分散의 65%를 要約하며 이것은 回歸分析의 R^2 과 똑같은 數學的 의미를 갖는다.

또한 各 要因에 의하여 요약된 各 變數의 분산을 커뮤니티(communality, h^2)라고 하며,

共通要因(common factor)들에 의하여 설명된 총분산의 %를 의미한다. 共通要因은 최소한 두개의 변수로 할당되어야 하며 그렇지 않은 모든 다른 要因들을 特別要因(specific factor)이라 한다. 따라서 한 변수의 총분산은 共通要因과 特別要因으로 나누어져 할당한다.

(表 1) 要因負荷行列

變 數	要 因		h ² (컴퓨터 리터)
	A	B	
VAR 001	.71	.40	.66
VAR 002	.70	.46	.70
VAR 003	.70	-.37	.63
VAR 004	.69	-.41	.64
VAR 005	.65	-.43	.61
VAR 006	.71	-.39	.66
自 乘 合 (eigenvalue) 分散要約 (eigenvalue/變數의 수)	2.89	1.01	3.90
	.48	.17	.65

表 1의 例에서 變數 VAR001에 대한 컴퓨터 리터(h₂)는 (.71)₂ + (.40)₂ = .66이다. 즉 이것은 VAR001의 총분산(여기서에는 1)의 66%를 要因 A와 B가 설명하고 있다.

各 要因負荷는 앞에서 설명한 바와 같이 변수와 要因間의 相關도를 나타낸 것이다. 따라서 우리는 各 要因의 要因負荷를 서로 곱하여 합하므로써 두변수의 상관관계를 얻을 수가 있다. 例를 두면 表 1에서 변수 VAR001과 VAR002의 相關도는 .68[(.71 × .70) + (.40 × .46)]이다.

2. 要因分析時 決定事項

要因分析을 하는데 數理的인 分析은 어떠한 方法을 사용하든지 비슷하나 주어진 資料에 맞는 方法을 선택하는 것은 매우 중요하다. 따라서 要因分析의 適當한 過程을 잘 선택하는 것은 기술이며 다음과 같은 사항들을 반드시 決定하여야 한다.

① 무엇에 무엇을 相關시켜 볼 것인가?

要因分析은 여러 관측된 사실로부터 關係를 찾아 보려는 것이다. 따라서 통상 入力으로부터 相關行列(correlation matrix)을 사용한다.

다. 만약 市場의 소비자 行態分析에서는 모든 資料들이 소비자와 이에 相關된 諸特性(attributes), 그리고 時間과 같은 三次元으로 집합될 수 있겠으며 모든 資料들은 이러한 3차원으로 설명된다. 마찬가지로 南北韓關係의 모든 事實들이 어떤 對象으로 설명될 수 있겠다 要因分析에서는 무엇에 무엇을 相關시키느냐에 따라 여러 형태의 方法이 있다. 즉

R型 要因分析: 가장 흔히 사용되는 方法으로 변수간의 相關關係(correlation among variables)分析에 使用.

Q型 要因分析: 個人 또는 對象物間의 상관관계분석에 使用.

O型 要因分析: 한 個人이나 對象의 時間에 대한 相關關係分析에 使用.

P型 要因分析: 한 個人이나 對象과 變數들간의 상관관계분석에 使用.

S型 要因分析: 한 변수에서 個人 또는 對象들의 상관관계분석에 使用.

따라서 分析家は 자기의 研究目的에 따라 위의 한 형태를 미리 결정하여야 한다.

② 入力으로써 相關行列(correlation matrix), 共通分散行列(covariance matrix), 참고자료 行列(cross-data-matrix)중 어느것을 택할 것인가?

흔히 要因分析에서는 相關行列을 사용하나 반드시 그렇게 하여야 하는 법은 없다. 相關係數는 모든 변수들의 平均을 0으로, 分散을 1로 놓고 標準點數(standard score)로 부터 구한 값이다. 이것은 무엇을 말하는가 하면, 資料行列에 포함된 3가지 형태의 情報중에서 변수들의 水準과 分散(dispersion)을 잃어버린것을 의미한다. 問題에 따라 이렇게 잃어버린 情報들을 必要로 할때가 있으므로 이럴때에 우리는 共通分散行列(covariance matrix) 또는 참고자료 行列(cross-data matrix)이 要因分析의 入力으로 適當하다. 一般的으로 測定된 단위가 변수들간에 차이가 심할때는 資料를 標準化해야 하므로 相關行列을 사용하는 것이 좋고 측정의 단위가 같거나 거의 비슷하고 個人的 差異가 豫測될때는 共通分散行列이나 참고자료 行列을 사용하는 것이 바람직하다. 그러나

要因分析에서 入力으로 相關行列을 주로 사용하는 이유는 要因負荷들이 변수간의 상관관계로 해석될 수 있고 評價될 수 있으며, 要因들이 -1과 +1사이의 값을 갖기 때문이다.

③ 相關行列의 대각선의 값은 무엇으로 할 것인가?

要因分析의 入力으로 相關行列을 사용한다고 決定했을때 하나 더 결정할 사항은 變數自身間的 相關도를 나타내는 行列의 對角線에 어떤 값을 넣느냐하는 것이다. 統計的인 意味에서 본다면 自己相關도는 1이어야 하므로 대각선의 값은 1을 넣어야 한다. 그러나 측정이 反復될때 측정오차때문에 똑같은 변수에도 完全한 상관을 얻기가 어렵기 때문에 信賴도를 고려하여 1보다 적은 값을 대각선에 넣기도 하며 특히 중요한 이유는 변수의 모든 分散이 共通要因에 의하여 모두 要約되지 않을 수가 있다는 것이다.

즉, 대각선에 1을 넣는 것은 主構成要素分析(principal component analysis)에 의하여 分散이 共通要因으로 분할된다는 가정때문이다. 그러나 特別要因들도 存在할 수 있기 때문에 대각선에 1이 아닌 다른값을 넣을 수도 있다는 것이다. 그러나 통상 이러한 경우 이론적인 뒷받침이 아직 확고하지 않아 대각선에 1을 넣어 사용하고 있다.

④ 회전(rotation)

다음에 決定해야 할 중요한 사항은 회전에 의하여 初期 結果로 부터 새로운 要因을 찾아내는 것이다. 옷감에 染色시 색깔을 달리 함에 따라 다른 모양을 나타내는 것과 같이 基本構造는 변하지 않더라도 回轉을 달리 함으로써 자료의 다른 構造를 볼 수가 있다. 이러한 회전은 要因分析에서 사용되는 단순한 構造의 原理를 이용한 것뿐이다. 즉 이 원리는 일련의 分析的 회전 選擇方法에 따라 數值化된다. 回轉하는 方法에는 여러가지가 있으나 통상 varimax회전이 널리 사용된다. 回轉方法이 다르면 결과도 完全히 달라진다. 統計的관점에서 보면, 사실상 모든 결과는 같지만 要因分析의 결과를 해석하기 위한 측면에서 살펴보면 회전은 극히 중요하다.

表 1의 例에서 6개 변수에 대한 要因分析의 결과가 回轉안된 要因負荷와 回轉된 要因負荷가 表 2에 있다. 앞에서도 여러번 언급했지만 要因分析의 主目的은 여러 변수들을 몇개의 共通要因으로 줄이는 것이다. 表 2에서 회전이 안된 要因負荷로 부터는 해석하기가 애매모호할 뿐더러 구분핑하기가 힘들다, 이것을 회전시키고 보니까 무언가 꼭 의미있는 결과를 얻은 것처럼 보인다. 즉 要因 A는 VAR004, VAR005, VAR006을 함께 표현할 수 있

(表 2) 回轉된 要因分析

變 數	非 回 轉			回 轉		
	A	B	h ²	A	B	h ²
VAR001	.71	.40	.66	.23	.78	.66
VAR002	.70	.46	.70	.18	.82	.70
VAR003	.70	.37	.63	.25	.75	.53
VAR004	.69	-.41	.64	.78	.18	.64
VAR005	.65	-.43	.61	.77	.14	.61
VAR006	.71	-.39	.66	.78	.21	.66
自 乘 合 (eigenvalue)	2.87	1.01	3.90	1.98	1.92	3.90
分散要約	.48	.17	.65	.33	.32	.65

는 것이며, 要因 B는 VAR001, VAR002, VAR003를 함께 표현하는 것이라고 볼 수 있다. 다시말하면 VAR004, VAR005, VAR006은 要因 A로, VAR001, VAR002, VAR003은 要因 B로 나타낼 수 있는 것이다. 이와 같이 回轉된 要因負荷는 자료의 構造를 分析하고 變數들을 구분화하는데 아주 좋다.

例에서 나타난 바와 같이 두 要因에 의하여 요약된 總分散은 회전하기 전이나 회전한 후나 변함없이 같으나 각 요인에 의하여 요약된 양은 늘 변한다. 또한 각 변수들의 communality(communality)는 변하지 않으나 두 요인에 의하여 분배된 양은 역시 변한다. 요약하면 회전할때 일어나는 현상은 要因들의 상관관계와 變數들의 內部的인 移動만 있을 뿐이고 外部的으로는 아무런 변화가 없다.

⑤ 언제 要因選定(factoring)을 멈추어야 하는가?

多數의 변수들을 要因別로 선정해야 할때

몇개의 요인으로 모든 변수들을 설명할 수 있을지 꼭 의문이다. 통상 가장 흥미있고 큰 변수들을 먼저 뽑아 내고 차차 작은 것으로 좁혀간다. 그러나 順次的으로 한 변수씩 要因化할때는 不必要하게 컴퓨터 시간을 낭비할 뿐만 아니라 回轉에 영향을 주기 때문에 변수의 해석을 오히려 애매하게 만들어 주는 경우가 있다. 많은 要因들이 varimax 回轉에 의할때 통상 몇개의 변수에 높은 要因負荷가 작동된 回轉된 要因들을 찾아내 준다. 적절하게 요인들을 선정하는 것은 초점을 맞추는 것과 똑같다고 볼 수 있으며, 너무 많은 요인들로 factoring할때는 오히려 의미를 흐리게 한다.

要因選定을 멈추는 시기는 대강 다음과 같은 4가지 尺度에 의한다 첫째 분석가가 미리 要因의 수를 정하여 미리 정해진 수만큼의 要因이 선정됐을때 멈추는 것이다. 둘째는 분석가가 미리 要因들에 의하여 설명될 수 있는 分散量(variance amount)을 알고 있을때, 이분산량에 맞는 尺度에 맞는 수의 要因을 선정한다. 흔히들 자료를 처리할때 위와 같은 사실들을 잘 모르므로 보통 요인들이 적은 要因負荷를 가지고 의미가 없어질때까지 요인선정을 계속한다.

셋째는 分散의 增加法이다. 즉 예를 들어 첫 번째 要因選定에서 전체 分散의 75%를 설명할 수 있게 나타났고 그 다음 하나의 要因을 더 추가했을때 5% 보다 더 적게 分散이 설명될때 통상 要因選定을 中止한다. 넷째는 相關行列이 要因分析될때 eigenvalue가 1보다 큰 모든 要因들을 중요하고 의미있는 要因으로 뽑아내는 方法이다.

지금까지 相關시켜야 할 對象, 入力の 形態, 相關行列, 대각선의 값, 사용한 回轉方法, 그리고 要因選定の 中止 尺度에 관한 분석가가 결정해야 할 사항들에 대하여 서술하였으며; 이러한 결정들은 여러가지 數值的인 結合을만 들어 내므로 그렇게 쉽지가 않다.

要因分析에 의하여 不必要한 변수를 제거하고 설명될 수 있는 변수만으로 回歸分析을 할 때 훨씬 정확한 계산을 얻을 수가 있다. 즉 통상 회귀분석을 하기 위하여 관계되는 모든 변

수들을 獨立變數로 하여 관계식을 찾아낸다. 이때 각 변수간의 상관관계가 存在할 수 있으므로 誤差項(error term)의 값이 커지게 된다. 따라서 回歸分析前에 要因分析을 하여 변수들간의 독립성을 검사하는 것은 반드시 必要하다.

要因分析을 위한 컴퓨터 프로그램은 여러 형태의 컴퓨터 팩키지로 쉽게 가용하며 특히 SPSS는 社會科學 分野의 問題들을 分析하는 아주 훌륭한 統計 팩키지⁽²⁾이며 韓國科學技術 研究所(KIST)의 CDC 6400에 설치되어 있다

Ⅲ. 要因分析의 應用

要因分析은 北韓의 行態를 豫測할 수 있는 적절한 方法의 하나이며 여기에서는 실제 자료를 직접 사용하여 어떤 결과를 얻는것보다는 南北韓關係의 어떤 問題들을 要因分析에 의하여 推定할 수 있는가하는 문제를 서술하고 大略的인 要因分析의 사용과정을 알아 보 고자 한다.

예를 들면 北韓의 經濟成長, 人口, 生産量 등의 經濟的인 豫測들은 현재 存在하는 經濟-經營問題에 사용되는 여러 豫測모델들을 應用할 수가 있다. 그러나 다분히 政治的인 복잡한 人間의 政策要因들이 混合된 국제政治의 未來狀況이나 國防問題를 豫測하다는 것은 計量的으로 심히 어려우며, 흔히 델파이법을 많이 사용한다. 그러나 要因分析은 北韓의 과거 형태를 기본으로 하여 앞으로의 北韓의 政策 方向에 대한 挑發을 豫測하는데 타당한 計量 分析이다.

우선 一般的으로 要因分析에 의하여 연구될 主題와 어떤 豫測 結果를 얻을 수 있는가를 알아본다.

北韓은 지난 6.25사변이후 수없이 많은 挑發行爲를 저질러 왔고 앞으로도 비슷한 도발이 있을 수 있겠다. 그렇다면 이러한 과거의 도발 사건들을 전부 나열하고 일정한 조사형

(2) Nie, Norman. H., etc., Statistical Package for the Social Sciences, 2nd ed., McGraw Hill Book Co., 1975

식에 의하여 그 잔인정도, 납치된 人員數, 침투경로, 도발방법, 北韓의 國內事情에 대한 問題, 南韓의 國內事情, 그리고 其他 美國, 日本, 中共, 소련의 태도등을 중심으로한 內容들을 도발사건별로 타당한 기술방식에 따라 자료를 수집한다.

各 挑發事件을 中心으로 조사한 자료를 미리 정한 변수들에 의하여 要因分析을 한다. 要因分析의 결과 몇개의 要因들로 변수들이 구분화될 것이며, 더 나아가 要因點數(factor score)에 의하여 各 挑發事件들을 要因別로 그 強度에 따라 分類할 수 있다. 이렇게 要因點數에 의하여 도발사건들이 구분화되면, 이러한 도발사건들은 北韓의 사정이 어떠한고 南韓의 사정이 어떠한때 일어났는가를 回歸分析에 의하여 모델화할 수가 있다. 이러한 회귀예측모델은 각 요인마다 얻어 질 수 있겠으며 이렇게 얻어진 모델을 통하여 현재 北韓과 南韓의 사정과 國際적인 여건을 고려하여 보면 北韓이 어떤 형태의 도발이나 정책을 취할 것인가를 豫測할 수 있는 것이다.

위의 일반적인 서술을 좀 더 자세히 問題化하면 다음과 같다.

① 6.25사변이후의 挑發事件들을 열거한다 예를 들면,

- case 1 : 미류나무사건
- case 2 : 김신조사건
- case 3 : 문세광사건
- case 4 : 프레블로호사건
- case 5 : 휴전선총격사건
- case 6 : 어선남북
- case 7 : 무장공비
- ⋮
- ⋮

위의 도발사건들과 같이 新聞이나 정보기관을 中心으로 수없이 많은 크고 작은 도발사건들을 찾을 수가 있다.

② 다음에 위에서 열거한 각 도발사건들을 미리 정한 변수에 따라 計量化한다. 變數의 選定은 정치학자 및 전문가에 의하여 적절하게 選定되어야 한다. 예를 들면,

- VAR001 = 殘認性
- VAR002 = 挑發方法

- VAR003 = 殺傷者數
- VAR004 = 對美國挑發
- VAR005 = 北韓內部權力狀態
- VAR006 = 北韓政治的 問題
- VAR007 = 北韓經濟問題
- VAR008 = 南韓經濟的 問題
- VAR009 = 南韓政治的 問題
- VAR010 = 南韓社會的 不安
- VAR011 = 美國의 對南韓政策
- ⋮
- ⋮
- VAR050 = 소련의 對北韓政策

위와같이 미리 선정한 변수에 따라 각 挑發事件을 몇 사람의 전문가가 數值化하여 要因分析의 初期 入力(input)으로 한다.

③ 컴퓨터 프로그램에 맞게 카-드를 작성하여 run한다. SPSS의 要因分析 프로그램을 사용했을때 얻어지는 output는 다음과 같은 것들이다.

㉔ 各 變數들의 平均 및 標準備差

VARIABLE	MEAN	STANDARD DEV	CASES
VAR001	.2418	.4287	7
VAR002	.3854	.4873	7
⋮	⋮	⋮	⋮
VAR050	5.5743	4.1001	7

여기에서 cases는 各 變數들의 자료의 數이며 挑發事件의 數와 같다.

㉕ 가장 基本이 되고 중요한 output는 相關係數(correlation coefficient)行列이다. 이 相關係數行列은 要因分析의 入力이 된다. 相關係數行列은 表 3과 같이 프린트된다.

㉖ 다음에 프린트되는 output는 要因을 몇 개 선정해야 할 것인가를 決定하는데 필요한 情報를 제공하는 것이며 表 4와 같다. 要因選定을 결정해야 할 尺度에 대하여 앞에서 說明했지만 表 4에 의하면 VAR001은 總分散의 15.7%를 설명하며, 만약 要因을 11개로 선정했을때는 總分散의 94.3%를 說明한다. 總分散의 90%(또는 경우에 따라 80%이상) 정도 설명해 주면 要因選定을 中止한다.

㉗ SPSS의 統制카-드 및 어떤方法的 要因

〈表 3〉 相關係數行列(Correlation Coefficient Matrix)

	VAR001	VAR002	VAR003	VAR049	VAR050
VAR001	1.000	0.3625	0.0402	0.1132	-0.047
VAR002	-0.3625	1.0000	0.0157	-0.0156	0.0316
VAR003	0.2117	0.0157	1.0000	0.0102	0.0438
⋮	⋮	⋮	⋮	⋮	⋮	⋮
VAR049	-0.1132	-0.0156	0.0899	1.0000	0.0436
VAR050	-0.037	0.0316	-0.1729	0.0436	1.0000

〈表 4〉 要因選定情報 output

VARIABLE	EST. COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
VAR001	1.000	1	3.142	15.7	15.7
VAR002	1.000	2	2.542	12.7	28.4
⋮	⋮	⋮	⋮	⋮	⋮
VAR010	1.000	10	7.519	3.8	91.9
VAR011	1.000	11	7.3335	3.7	94.3
⋮	⋮	⋮	⋮	⋮	⋮
VAR050	1.000	50	3.268	1.6	100.0

〈表 5〉 要因負荷行列(Factor Matrix Using Principal Component)

	FACTOR1	FACTOR2	FACTOR3	FACTOR11
VAR001	-.6926	.3782	-.089700119
VAR002	.4826	-.2421	.140104556
VAR003	.0232	-.1367	.12680273
⋮	⋮	⋮	⋮	⋮	⋮
VAR050	.2062	.3950	-.1987	-0.3792

分析을 쓰는가의 選擇코드에 따라 output의 내용이 달라진다. 主構成要素法 (principal component)에 의하여 各 變數의 要因負荷行列(factor loading matrix)이 주어진다. 다시 말하면 앞의 相關係數行列로 부터 要因負荷가 계산되며 프린트된다. 表 5는 各 挑發事件의 變數들에 대한 要因負荷를 나타낸 것이나 통상 이것으로 부터는 分명한 해석을 내리기가 어렵다. 따라서 回轉(rotation)이 必要하게 된다.

⊕ VARIMAX나 QUARTIMAX 또는 EQUIMAX에 의하여 回轉시켜 해석하기 쉽도록 한다. 이 output는 要因分析에서는 가장 중요하고 받드시 얻어야 하는 부분이다. 이것으로부터 우리는 몇개의 變數들을 한개의 要因으로 묶고 어떤 意味를 이 要因에 부여한다. 다시 말하면 表 5에서는 도저히 變數들을 要因別로 구분화하기가 어렵고 어떻게 要因을 설명해야 할지 몰랐으나 回轉을 시켜 놓고 보면 變數가 要因負荷值에 의하여 구분화된다.

VARIMAX에 의하여 回轉시켰을때 얻어지는 output는 表 6의 回轉된 要因負荷行列이다 表 6에서 VAR001과 VAR003은 FACTOR1에 대하여 높은 要因負荷를 가지고 있으므로 FACTOR1로 要約할 수가 있다. 즉 VAR001은 殘忍性, VAR003은 殺傷者數에 대한 변수들이므로 FACTOR1에 의미를 부친다면 殘忍程度를 나타내는 것이라고 볼 수 있으며 VAR001과 VAR003을 하나의 변수로 취급할 수 있는 것이다.

〈表 6〉 回轉要因負荷行列
(VARIMAX ROTATED FACTOR HATPIX)

	FACTOR1	FACTOR2	...	FACTOR11
VAR001	.7325	.35210131
VAR002	.2453	.43200010
VAR003	.8131	.21300531
⋮	⋮	⋮	⋮	⋮
VAR009	.1124	.79331401
VAR010	.0734	.68110089
⋮	⋮	⋮	⋮	⋮
VAR050	.3254	.45310287

또한 VAR009와 VAR010을 FACTOR2로 요약되며 VAR009는 南韓政治의 狀態를 VAR010은 南韓의 社會的 不完을 표시하는 변수들이므로 FACTOR2는 南韓의 政治的 不安定을 나타내는 것이라고 볼 수 있겠다. 물론 여기에서는 要因分析을 南北韓關係豫測에 사용하는 方法을 설명하기 위한 것이므로 가상 data를 가지고 설명하고 있으나 表 6의 要因負荷行列은 各 變數들을 要約하고 FACTOR들의 성질을 identify하고 해석하는 가장 중요한 것이다.

回歸分析을 위하여 변수들을 선정하기 위하여서는 통상 回轉要因負荷行列을 얻으면 된다. 그러나 각 道발사건들을 좀 더 分析하고 豫測을 위하여서는 要因點數(factor score)를 구하여야 한다. 各挑發事件(case)의 要因點數는 各 case의 各 변수에 대한 값과 要因點數係數(factor score coefficient)를 곱한 값이며 이 要因點數(factor score)에 의하여 各 要因別로 挑

發事件들을 歸화할 수 있는 것이다.

表 7은 要因點數係數를 나타낸 것이며 表 8은 要因分析의 最終 output로 要因點數(factor score)를 나타낸 것이다.

case *i*의 FACTOR *j*에 대한 factor score는 다음과 같이 구하여 진다.

$$\text{Factor Score} = (\text{VAR001에 대한 factor score coeff.}) \text{VAR001} + (\text{VAR002에 대한 factor score coeff.}) \text{VAR002} + \dots +$$

$$(\text{VAR050에 대한 factor score coeff.}) \text{VAR050}$$

〈表 7〉 要因點數係數
(Factor Score Coefficient)

	FACTOR1	FACTOR2	FACTOR11
VAR001	-.2641	-.00361609
VAR002	.1502	.05241231
⋮	⋮	⋮	⋮	⋮
VAR050	.0453	-.04790458

〈表 8〉 要因點數(Factor Score)

挑發事件	FACTOR1	FACTOR2	FACTOR11
case1	-.3564	.4042	1.092
case2	-.3942	.50244809
case3	-.1951	.1234	-1.453
⋮	⋮	⋮	⋮	⋮

FACTOR1은 殘忍性程度를 나타내는 要因이며 要因點數에 의하면 8.18미루나무사건과 金心조사건이 殘忍性에 대하여 같은 比重을 가지고 있다. 이렇게 要因點數를 가지고 6.25이후 北韓이 저질렀던 道발사건들을 各 要因別로 區分할 수 있겠고 大略의 變數 要因과 道발사건과의 關係를 推定할 수가 있다.

北韓의 未來狀況을 豫測하기 위하여서는 結論을 더 나아가 要因點數에 의한 結果를 더 擴大 分析할 必要가 있다. 즉 FACTOR1로 몇 개의 道발사건들이 歸화되었을때 이러한 事件들을 中心으로 다시 要因分析을 하고 回歸分析을 하면 南韓과 北韓의 國內外 政治 經濟

社會與件이 어떻고 美國을 비롯한 주변 四強國의 관계가 어떠할때 어떠한 도발사건이 일어났다는 豫測모델을 끌어낼 수가 있고 이 豫測모델을 가지고 현재 南韓의 實情, 北韓의 現況 그리고 美國, 中共等の 政策들을 연구하면 北韓이 앞으로 어떻게 어떤 형태의 도발을 일으킬 것인가를 豫測할 수가 있는 것이다.

IV. 結 論

아무리 不正確한 豫測 結果라도 豫測을 하지 않는 것보다는 훨씬 좋은 것이다. 따라서 資料가 充分하지 못하고 豫測모델이 完全치 못하더라도 北韓에 대한 information file을 조속히 完成하여 豫測을 하는 환경속에서 北韓

을 研究하고 分析하여 꾸준히 南北韓關係豫測에 적합한 모델을 개발하고 수정하여야 한다.

定性的인 豫測分析은 델파이법에 의하여 수행하고 이와 함께 시뮬레이션의 應用도 바람직하다. 앞에서 제시한 要因分析(factor analysis)에 의한 北韓의 앞으로의 挑發時期, 挑發方法, 挑發強度等の 行態豫測은 적합한 方法이라고 생각되며 이를 위한 資料가 잘 수집되고 要因分析에 의한 豫測 結果가 도출된다면, 南韓의 現정세를 판단함으로써 쉽게 여기에 대한 對應策을 마련할 수 있겠다.

물론 더 精確한 豫測을 위하여 定性的인 豫測모델은 물론, 定量的 豫測모델들이 要因分析과 함께 應用되어야 하고 信賴度와 精確성을 높이는 相互補完作用을 해야 할 것이다.