

Chapter08

주성분과 요인 분석

I 요인 분석

M T W T F S S

2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

FASTCAMPUS
ONLINE

금융공학/퀀트 I

강사. 장순용

I 키포인트

- 요인 분석 (Factor Analysis, FA).
- 주성분과 요인의 비교.
- 직교회전과 사교회전.

I 요인 분석 (FA) & 주성분 분석 (PCA)

- 먼저 X_1, X_2, \dots, X_p 와 같이 p 개의 변수가 있다고 가정한다.
- 주성분 분석으로 p 개의 주성분 PC_1, PC_2, \dots, PC_p 을 얻을 수 있다.

⇒ 개개 주성분은 원 변수의 선형 조합이다.

$$PC_i = \alpha_{1,i}X_1 + \alpha_{2,i}X_2 + \dots + \alpha_{p,i}X_p$$

⇒ 반대로 원 변수는 주성분의 선형 조합으로 표현할 수 있다.

$$X_i = \beta_{1,i}PC_1 + \beta_{2,i}PC_2 + \dots + \beta_{p,i}PC_p$$

I 요인 분석 (FA) & 주성분 분석 (PCA)

⇒ 차원축소할 때에는 적은 수 ($q < p$) 의 주성분만을 사용한다.

$$X_i \approx \beta_{1,i}PC_1 + \beta_{2,i}PC_2 + \cdots + \beta_{q,i}PC_q$$

⇒ 주성분은 분산의 크기로 정렬되어 있으니 첫 몇개만 사용한다.

$$\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \cdots$$

or

$$\sigma_1 > \sigma_2 > \sigma_3 > \cdots$$

I 요인 분석 (FA) & 주성분 분석 (PCA)

- 요인 분석도 적의 수의 **잠재 요인**으로 원 변수를 설명하는 것이다.

$$X_i = \beta_{0,i} + \beta_{1,i}F_1 + \beta_{2,i}F_2 + \cdots + \beta_{q,i}F_q + \varepsilon_i$$

⇒ F_1, F_2, \dots, F_q 는 **공통적**으로 나타나는 **잠재 요인**.

⇒ 주성분과는 달리 개수 (q) 를 먼저 정해놓고 구한다.

⇒ ε_i 는 변수 X_i 의 개별 오차이며 F_1, F_2, \dots, F_q 와는 독립적이다.

⇒ 가중치 $\beta_{1,i}, \beta_{2,i}, \dots, \beta_{q,i}$ 를 loading (**요인 부하량**) 이라 부른다.

⇒ 요인으로 설명 가능한 변수의 변동 비율을 communality (**공통성**) 이라

I 요인 분석 (FA) & 주성분 분석 (PCA) : 유사성

- 다음과 같은 유사성이 있다.

⇒ 주성분 (PC) \approx 요인 (factor).

⇒ 차원축소를 목표로 한다. 차원 축소 후 주성분의 개수 \approx 요인의 개수.

⇒ 원 변수 사이의 공유하는 구조 발견을 목적으로 한다.

I 요인 분석 (FA) & 주성분 분석 (PCA) : 차이점

- 다음과 같은 차이점이 있다.

요인에는 **주관적**인 의미 부여 \Leftrightarrow 주성분은 객관적 해석 우선
시

요인은 다 **동등한** 취급을 받는다 \Leftrightarrow 주성분은 분산의 크기로 차
등을 둠

요인의 개수는 분석 **전**에 정해둔다 \Leftrightarrow 주성분의 개수 = 원 변수의
개수

I 요인 분석 (FA) & 주성분 분석 (PCA) : 차이점

- 다음과 같은 차이점이 있다.

요인에는 **주관적**인 의미 부여 \Leftrightarrow 주성분은 객관적 해석 우선
시

요인은 다 **동등한** 취급을 받는다 \Leftrightarrow 주성분은 분산의 크기로 차
등을 둬
요인 ~~의 수~~ ^{주관적 또는 상관계} ~~를~~ ^수 ~~정해둔다~~ ^를 ~~정해둔다~~ \Leftrightarrow 주성분의 개수 = 원 변수의
개수 ^{기를 분석하여 정한} 다.

I 요인 분석 (FA) & 주성분 분석 (PCA) : 차이점

- 다음과 같은 차이점이 있다.

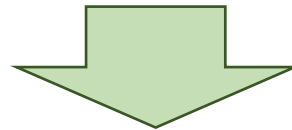
요인에는 **주관적**인 의미 부여 \Leftrightarrow 주성분은 객관적 해석 우선
시

요인은 다 **동등한** 취급을 받는다 \Leftrightarrow 주성분은 분산의 크기로 차
등을 둠

요인의 개수는 분석 **전**에 정해둔다 \Leftrightarrow 주성분의 개수는 누적분산
비를 산출해서 정한다. 변수의

I 요인 분석 (FA) & 주성분 분석 (PCA) : 관점에서의 차이점

X_1 X_2 X_3 ... X_p



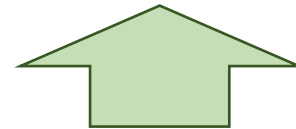
×가중치

PC_1 PC_2 PC_3 ... PC_p

PCA는 원 변수의 선형 조합으로 주성분 계산을 우선시.

I 요인 분석 (FA) & 주성분 분석 (PCA) : 관점에서의 차이점

X_1 X_2 X_3 ... X_p



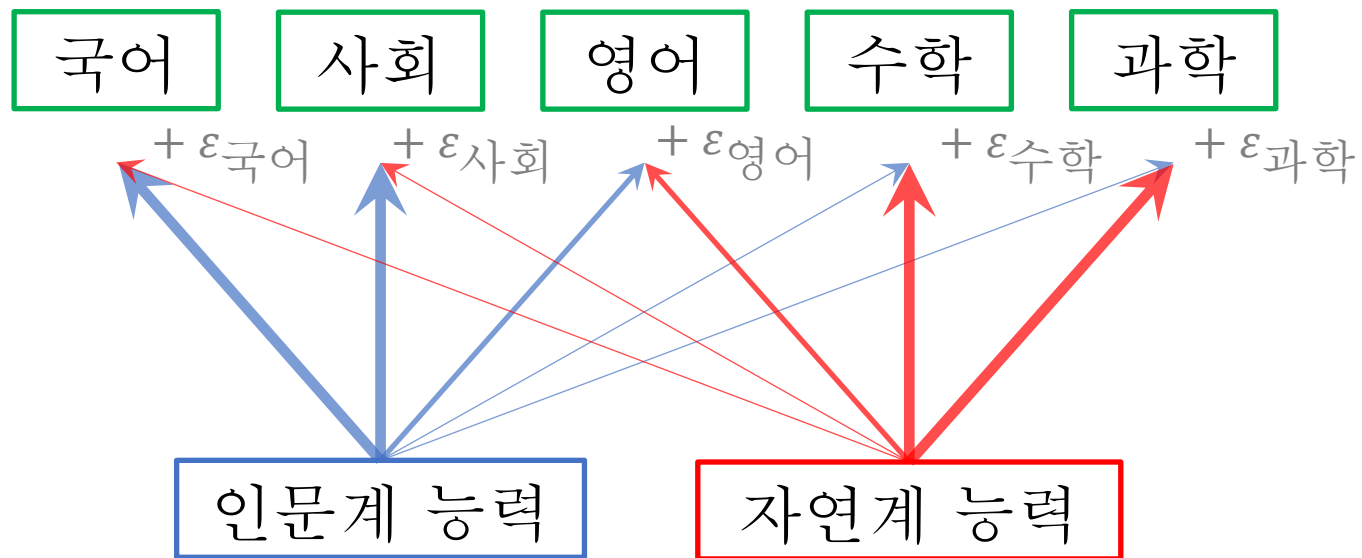
\times 가중치 $+ \varepsilon_i$

F_1 F_2 ... F_q

FA는 잠재 요인을 발굴해서 원 변수의 **주관적** 설명을 우선시.

I 요인 분석 (FA)의 예

- 다섯 개의 관측 변수를 설명하는 두 개의 요인.



I 회 전

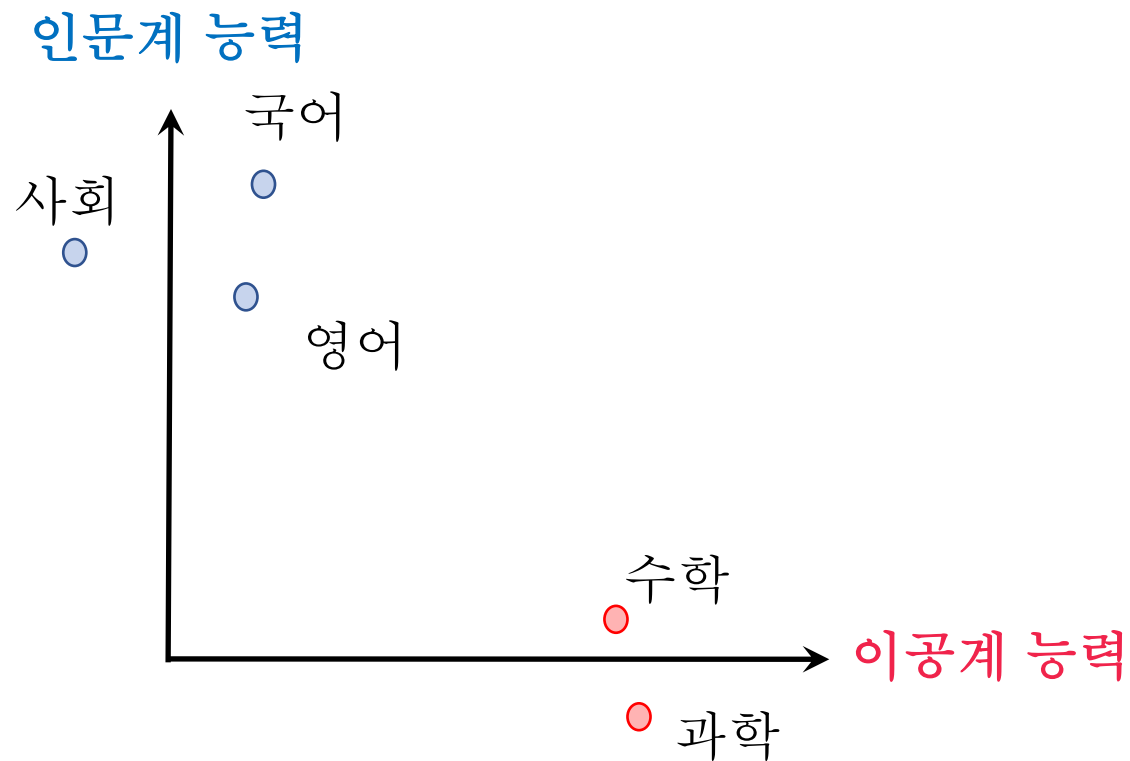
- 요인 부하량을 가지고 요인의 주관적 해석과 naming을 한다.

$$X_i = \beta_{0,i} + \beta_{1,i}F_1 + \beta_{2,i}F_2 + \cdots + \beta_{q,i}F_q + \varepsilon_i$$

예). 국어 $\cong 0.9 F_1 + 0.1 F_2$
 과학 $\cong 0.1 F_1 + 0.9 F_2$ } F_1 는 “인문계 능력”
 F_2 는 “이공계 능력”이
 라 해석하고 naming한
 다.

I 회 전

- 요인을 좌표축으로 정하고 부하량을 사용하여 원 변수를 plot할 수 있다.

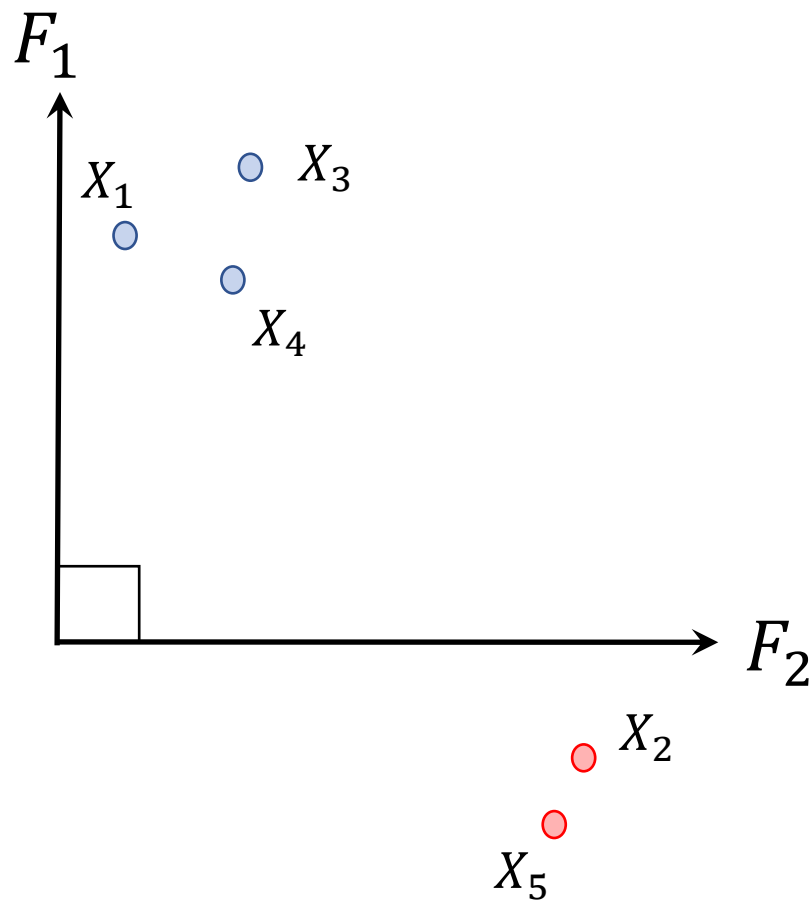


I 회전

- 그런데 대다수의 경우 최적화된 해석을 위해서 좌표축의 회전이 필요하게 된다.
 - ⇒ 직교회전 : 좌표축의 직교관계 유지. Varimax와 같은 방법.
 - ⇒ 사교회전 : 좌표축이 별개로 회전. Promax, Oblimin, 등의 방법.

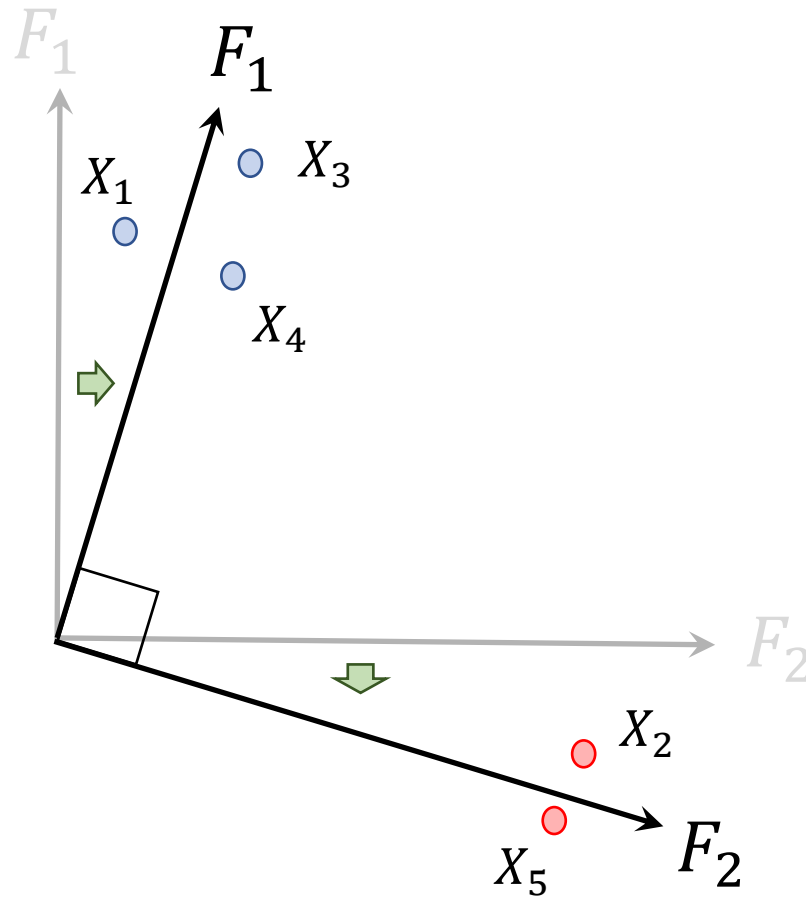
I 직교회전 (orthogonal rotation)의 원리

- 좌표축 (요인)의 방향이 최적화 되어 있지 않다.



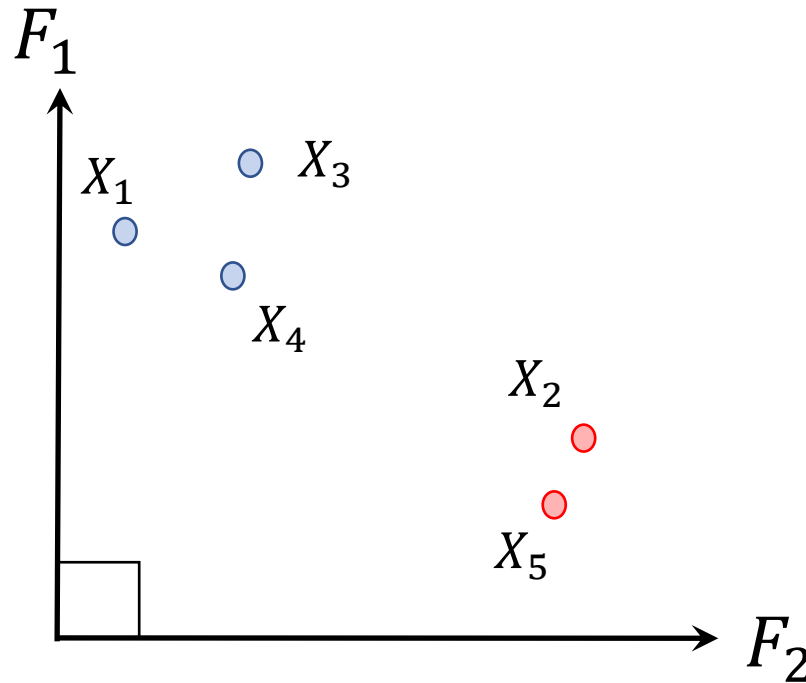
I 직교회전 (orthogonal rotation)의 원리

- 직교회전 후 요인을 해석하고 naming하기 수월해 진다.



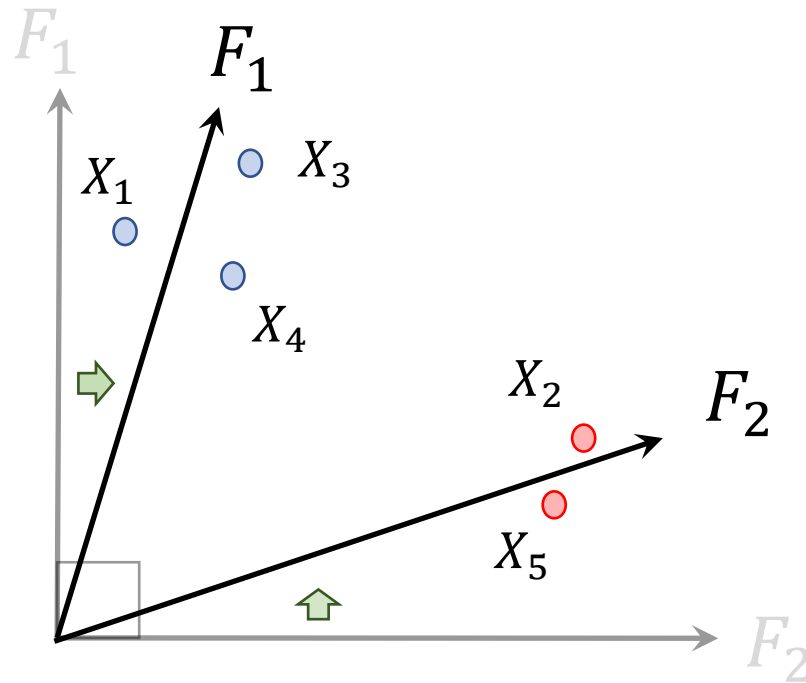
I 사교회전 (oblique rotation)의 원리

- 좌표축 (요인)의 방향이 최적화 되어 있지 않다.



I 사교회전 (oblique rotation)의 원리

- 사교회전 후 요인을 해석하고 naming하기 수월해 진다.



I 요인 분석의 순서

- R과 같은 통계 툴을 사용하여 쉽게 실행할 수 있다.

요인수 결정: 상관계수 행렬의 고유값, 등.



계산 실행: 최대 우도법, 최소 제곱법, 주성분, 등.

요인의 회전: 직교, 사교.



요인의 해석 및 naming.

I 끝.

감사합니다.

