# 9.1 9.2 정준변수와 정준상관

2019년 6월 16일 일요일 오전 3:09

두 개의 변수 집단 간의 관계를 파악하고 양으로 표현하고자 할 때 정준상관분석을 할 수 있다.

두 변수집단간의 선형적 연관성은 정준상관계수라는 측도로 표현됨.

(한 개 변수, 한 개 변수) 에 대한 단순상관계수와 (한 개 변수, 여러 개 변수) 에 대한 다중상관계수 개념의 일반화로 정준상관계수를 이해할 수 있음.

정준상관분석에서는 (여러 개 변수, 여러 개 변수) 에 대한 상관계수문제를 다룸.

정준상관분을 통해 종속변수들과 독립변수들의 두 집단 간의 정준상관계수를 계산할 수 있음. => 두 변수 집단간의 관계 파악

다차원에 놓인 두 변수 집단간의 관계를 저차원의 정준변수 쌍으로 전환하여 관계를 설명할 수 있음.

정준상관계수가 정준변수간의 상관성을 나타냄.

정준변수와 정준상관계수를 구하는 단계

1) 가장 큰 상관계수를 갖는 한 쌍의 선형결합식을 결정함.
2) 첫 번째로 선택된 한 쌍의 선형결합식과는 독립이면서 그 다음으로 큰 상관계수를 갖는 선형결합식을 찾음.
3) 이와 같은 방법으로 먼저 찾은 선형결합식들과는 독립이면서 그 다음으로 큰 상관계수를 갖는 선형결합식을 찾음.

이렇게 찾은 변수들의 선형결합식을 정준변수라고 하고 이들의 상관계수를 정준상관계수라고 함.

정준상관계수는 두 변수 집단간의 연관성 정도를 나타냄.

정준상관계수는 다중상관계수의 확장으로 볼 수 있음.

즉, 한 개의 변수와 여러 개 변수간의 상관관계를 여러 개 변수와 여러 개 변수간의 상관성으로 확장한 개념.

### 9.2.1 정준변수와 정준상관계수의 정의 및 개념

다중상관계수를 구하는 방법 (p238)

(여러 개 변수, 여러 개 변수) 에 대해 두 변수 집단간의 상관성을 나타내는 방법

각 개체에 대해 두 개의 변수 집단 $X = (X_1, \cdots, X_p)'$ 와 $Y = (Y_1, \cdots, Y_q)'$ 가 관측되었음.

두 변수 집단으로 구성된 $(p+q)\times 1$ 확률벡터 W는 다음과 같이 표현됨.

$$W = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \\ \hline Y_1 \\ \vdots \\ Y_q \end{pmatrix} \qquad E(W) = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \qquad \Sigma_{(p+q)\times(p+q)} = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}$$

변수들의 선형결합식을 다음과 같이 표현함.

상수계수벡터로서 $p\times 1$ 벡터 a와 $q\times 1$ 벡터 b에 대해 U, V는 일변량 확률변수가 됨.

$$U = a'X \qquad V = b'Y$$

$$Var(U) = a'Cov(x)a = a'\Sigma_{xx}a \qquad Var(V) = b'Cov(Y)b = b'\Sigma_{YY}b$$

$$Corr(U.V) = \frac{a'\Sigma_{XY}b}{\sqrt{a'\Sigma_{XX}a}\sqrt{b'\Sigma_{YY}b}}$$

$$Corr(U,V) = \frac{a'\Sigma_{xy}b}{\sqrt{a'\Sigma_{xx}a}\sqrt{b'\Sigma_{yy}b}}$$

U 와 V의상관계수를 최대로하는 상수벡터 a 와 b를 찾고자함.

정준변수를 구하는 과정

1) 첫번째 정준변수 쌍 $(U_1, V_1)$은 $Corr(U,V)$ 를 최대로 하며 $Var(U_1) = Var(V_1) = 1$ 인 변수들의 선형결합식이다.

2) 두 번째 정준변수 쌍 $(U_2, V_2)$ 는 $(U_1, V_1)$ 과 서로 독립이면서 $Corr(U,V)$ 를 최대로 하며 $Var(U_2) = Var(V_2) = 1$ 인 변수들의 선형결합식이다.

3) k번째 정준변수 쌍 $(U_k, V_k)$ 는 $(U_i, V_i)$ 과 서로 독립이면서 $Corr(U,V)$ 를 최대로 하며 $Var(U_k) = Var(V_k) = 1$ 인 변수들의 선형결합식이다.

정리9.1

$p \le q$ 라고 할때 $p\times 1$ 확률벡터 $X$ 와 $q\times 1$ 확률벡터 $Y$ 가

공분산행렬 $Cov(X) = \Sigma_{xx}$, $Cov(Y) = \Sigma_{yy}$ $Cov(X,Y) = \Sigma_{xy}$ 를 가질때

상수계수벡터로서 $a, b$ 와의 선형결합식 $U = a'X$, $V = b'Y$ 에 대해

최대상관계수 $\max_{a,b} Corr(U,V) = \rho_1^*$ 를 갖는 첫번째 정준변수는 $U_1 = e_1'\Sigma_{xx}^{-\frac{1}{2}}X$ 와 $V_1 = f_1'\Sigma_{yy}^{-\frac{1}{2}}Y$ 로 주어짐.

k번째 정준변수는 $U_k = e_k'\Sigma_{xx}^{-\frac{1}{2}}X$, $V_k = f_k'\Sigma_{yy}^{-\frac{1}{2}}Y$ 로 주어지며 $i = 1, \cdots, k-1$ 번째 정준변수와 서로 독립이면서

$Corr(U_k, V_k) = \rho_k^*$ 를 최대로 한다.

$\rho_1^{*2} \ge \cdots \ge \rho_p^{*2}$ 는 $\Sigma_{xx}^{-1/2}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1/2}$ 의 고유값이며 고유벡터는 $e_1, \cdots e_p$ 이다.

$\Sigma_{yy}^{-1/2}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1/2}$ 의 고유값을 순서대로 늘어놓을때 고유벡터는 $f_1 \cdots f_q$ 이다.

$f_i$ 는 $\Sigma_{yy}^{-1/2}\Sigma_{yx}\Sigma_{yx}\Sigma_{xx}^{-1/2}e_i$ 에 비례한다.

$Var(U_k) = Var(V_k) = 1$

$Cov(U_k, U_l) = Corr(U_k, U_l) = 0$

$\text{Var}(U_k) = \text{Var}(V_k) = 1$

$\text{Cov}(U_k, U_l) = \text{Corr}(U_k, U_l) = 0$

$\text{Cov}(V_k, V_l) = \text{Corr}(V_k, V_l) = 0$

$\text{Cov}(U_k, V_l) = \text{Corr}(U_k, V_l) = 0$

### 9.2.2 표준화변수에 대한 정준변수와 정준상관계수

표준화변수의 정준상관변수의 계수는 원래의 변수로부터 구한것에 $\sqrt{\sigma_{ii}} = \sqrt{\text{Var}(X_i)}$ 를 곱한형태로 얻어짐

$a_k'$ 가 k번째 정준변수 $U_k$ 의 계수벡터이면 표준화변수 Z의 계수벡터는 $a_k' D_{xx}^{1/2}$ 로 얻어짐

$D_{xx} = \text{diag}\{\sigma_{11}, \cdots \sigma_{pp}\}$ 변수들의 분산으로 구성된 대각행렬

$b_k' D_{YY}^{-1/2}$ 는 두번째 변수 집단 Y의 표준화변수에 대한 정준계수벡터가 됨.

표준화변수들에 대한 정준상관계수는 변하지 않는다.

9.3절 표본정준변수와 표본정준상관계수는 P244.

# 9.4 정준변수에 대한 해석 및 특성

2019년 6월 16일 일요일 오전 3:10

정준변수는 인공적으로 만들어낸 변수이므로 인자나 주성분과 같은 절대적의미를 부여하기 힘듦.

관심있는 변수 집단에 대해 연관성을 알고자 할 때 주로 이용할 수 있음.

그러나 변수를 표준화하더라도 정준상관계수는 변하지 않으므로 단위의 표준화와 해석을 위해서는 표준화변수들에 대한 정준상관분석을 권장함.

정준변수의 특성

1) 정준상관계수는 변수들의 척도변환에 불변이다.

2) 첫번째 정준상관계수 $\rho_1^*$ 는 두 변수 집단간의 최대 상관계수이며 두 변수 집단에서 단순상관계수 또는 다중상관계수를 구할 때 $\rho_1^*$ 를 넘지 않는다.

(1) 표준화된 계수

정준변수를 구성하는 정준계수는 정준상관계수에 각 변수가 기여하는 정도를 나타냄.

변수들의 단위를 통일하기 위하여 표준화변수들에 대해 구한 정준변수의 정준계수는 각 변수가 정준변수에 상대적으로 기여하는 바를 나타냄.

그러므로 변수들 중 일부가 제거되거나 첨가되면 상대적인 기여도가 변하므로 정준계수도 변하게됨.

정리 9.3 변수 Y와 정준변수 U와의 상관계수의 가중합 (p249)

(2) 각 변수와 정준변수와의 상관성 (p249)

# 9.5 상관성에 대한 검정

2019년 6월 16일 일요일 오전 3:10

정준상관계수와 관련된 검정.

정준상관분석은 두 변수 집단간의 연관성에 대해 설명하고자 하는것.

두 변수 집단간에 상관성이 존재할 때만 의미있는 분석이 됨.

상관성 존재 여부에 대한 검정이 필요함.

$H_0 : \Sigma_{XY} = O \qquad H_1 : \Sigma_{XY} \neq O$

귀무가설이 사실이면 X 와 Y 간에 연관성이 없으며 구해지는 정준상관계수 r 는 통계적인 의미가 없다.

검정통계량은 P250