



ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

PHÂN TÍCH NHÂN TỐ (Factor Analysis)

Giảng viên hướng dẫn: Lê Xuân Lý

Nhóm sinh viên: Nhóm 2

Học phần: MI4024

Mã lớp: 146157



Thành viên

Lê Thị Hằng - 20216924

Đào Ngọc Trinh - 20216963

Lê Ngọc Hà - 20216922

Nguyễn Thị Linh Chi - 20216913

Lê Đức Việt - 20216968

Vũ Danh Trung Hiếu - 20216925

Lê Anh Việt - 20216967

Nguyễn Mai Phương - 20216955

Hà Khánh Linh - 20216938

Nguyễn Trường Giang - 20216921

Vũ Quỳnh Anh - 20216912

Giới thiệu chung

Phân tích nhân tố là một công cụ mạnh mẽ dành cho các nhà nghiên cứu và nhà phân tích tìm cách hiểu các cấu trúc và mối quan hệ cơ bản trong dữ liệu.

Phân tích nhân tố cho phép các nhà nghiên cứu giảm độ phức tạp của một số lượng lớn các biến quan sát bằng cách xác định một tập hợp nhỏ hơn các yếu tố cơ bản nắm bắt thông tin thiết yếu có trong dữ liệu.

Phân tích nhân tố giúp xác định và khái niệm hóa các biến tiềm ẩn hoặc không thể quan sát có ảnh hưởng đến các biến quan sát.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng

Ý tưởng đầu tiên về phân tích nhân tố đã được nêu ra bởi K.Pearson và C.Spearman vào đầu thế kỷ XX để xác định và đo mức độ của trí thông minh.

Mục đích: Phân tích nhân tố là mô tả những mối quan hệ tương quan giữa nhiều biến thông qua ít biến, các biến này không thể quan sát được, được gọi là các yếu tố.

Mối quan hệ tương quan: Mối quan hệ tương quan giữa hai biến thường được biểu diễn qua hai khái niệm chính: Hiệp phương sai và Hệ số tương quan.

- Hiệp phương sai thể hiện mối quan hệ giữa hai biến với nhau, có thể là đồng biến hoặc nghịch biến. Nếu X, Y độc lập thì $\text{Cov}(X, Y) = 0$.
- Hệ số tương quan được sử dụng để đo lường mức độ quan hệ tuyến tính giữa 2 biến.

Đặt vấn đề

Giả sử tất cả các biến số trong một nhóm cụ thể có mối quan hệ tương quan cao với nhau, nhưng có các mối quan hệ tương quan tương đối nhỏ với các biến số trong một nhóm khác. Sau đó, có thể tưởng tượng rằng mỗi nhóm biến số đại diện cho một cấu trúc cơ bản duy nhất.

Ví dụ 1.1

Các mối quan hệ từ nhóm điểm kiểm tra trong các môn kinh điển: Tiếng Pháp, Tiếng Anh, Toán Học và Âm Nhạc được thu thập bởi Spearman gợi ý về một yếu tố "trí tuệ" cơ bản. Một nhóm biến số thứ hai, đại diện cho các điểm thể chất, nếu có sẵn, có thể tương ứng với một yếu tố khác.

Phân tích yếu tố để làm gì?

Ví dụ 1.2

Một trường đại học muốn đánh giá hiệu quả của các phương pháp giảng dạy trên nhiều môn học. Họ thu thập dữ liệu từ sinh viên bao gồm điểm số, tham gia lớp học, và đánh giá giảng viên. Sử dụng phân tích yếu tố, họ có thể xác định các yếu tố như chất lượng giảng viên, hỗ trợ học tập, và tương tác sinh viên - giáo viên. Thông qua thông tin này, trường đại học có thể tối ưu hóa các phương pháp giảng dạy để cải thiện trải nghiệm học tập của sinh viên.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao**
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng

Mô hình phân tích nhân tố trực giao

Xét vector ngẫu nhiên có thể quan sát được $X^T = (X_1, \dots, X_p)$ có $E(X) = \mu$, $cov(X) = \Sigma$. Mô hình nhân tố giả định rằng mỗi X_i là tổ hợp tuyến tính của một số ít biến ngẫu nhiên không quan sát được F_1, \dots, F_m (với $m < p$) gọi là các nhân tố chung và p biến cộng thêm $\epsilon_1, \dots, \epsilon_p$ được gọi là các sai số hoặc các nhân tố xác định.

Mô hình phân tích nhân tố trực giao

Ta có mô hình nhân tố trực giao:

$$\begin{aligned}X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \epsilon_1 \\X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \epsilon_2 \\&\vdots \\X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \epsilon_p\end{aligned}\tag{1}$$

hoặc dưới dạng ma trận:

$$\underset{(p \times 1)}{X} - \underset{(p \times 1)}{\mu} = \underset{(p \times m)}{L} \times \underset{(m \times 1)}{F} + \underset{(p \times 1)}{\epsilon}\tag{2}$$

Mô hình phân tích nhân tố trực giao

Trong đó, ma trận L có kích thước $p \times m$ và được gọi là ma trận hệ số tải nhân tố. Các phần tử của ma trận L là l_{ij} , là các hệ số tải.

Việc tìm F_1, \dots, F_k và $\epsilon_1, \dots, \epsilon_k$ là một bài toán không giải được. Tuy nhiên nếu thêm giả thiết:

$$E(F) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(m \times 1)} ; \quad \text{cov}(F) = E(FF^\top) = I_{(m \times m)}$$

$$E(\epsilon) = 0; \quad \text{cov}(\epsilon) = E(\epsilon\epsilon^\top) = \Psi_{p \times p} = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix}$$

Mô hình phân tích nhân tố trực giao

F và ε độc lập (không tương quan):

$$\text{cov}(\varepsilon, F) = E(\varepsilon F^\top) = 0_{(p \times m)}$$

thì bài toán sẽ có lời giải.

Mô hình phân tích nhân tố trực giao

Như vậy ta cần xét mô hình nhân tố trực giao dưới đây:

$$\underset{(p \times 1)}{X} - \underset{(p \times 1)}{\mu} = \underset{(p \times m)}{L} \times \underset{(m \times 1)}{F} + \underset{(p \times 1)}{\epsilon} \quad (3)$$

với $\mu_i = E(X_i)$

ϵ_i là nhân tố xác định thứ i

F_j là nhân tố chung thứ j

l_{ij} là tải trọng của biến X_i đặt lên nhân tố thứ j (Y_j)

$E(F) = 0; E(\epsilon) = 0$

$$\text{cov}(F) = I; \text{cov}(\epsilon) = \psi = \text{diag}(\psi_1, \dots, \psi_k) \quad (4)$$

$\text{cov}(\epsilon, F) = 0$

Mô hình phân tích nhân tố trực giao

Ta có:

$$\begin{aligned}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top &= (\mathbf{LF} + \boldsymbol{\varepsilon})(\mathbf{LF} + \boldsymbol{\varepsilon})^\top \\&= (\mathbf{LF} + \boldsymbol{\varepsilon})((\mathbf{LF})^\top + \boldsymbol{\varepsilon}^\top) \\&= \mathbf{LF}(\mathbf{LF})^\top + \boldsymbol{\varepsilon}(\mathbf{LF})^\top + \mathbf{LF}\boldsymbol{\varepsilon}^\top + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \\ \Sigma = \text{cov}(\mathbf{X}) &= E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top \\&= \mathbf{LE}(\mathbf{FF}^\top)\mathbf{L}^\top + \mathbf{LE}(\mathbf{F}\boldsymbol{\varepsilon}^\top) + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) \\&= \mathbf{LL}^\top + \boldsymbol{\Psi}\end{aligned}$$

Mô hình phân tích nhân tố trực giao

Suy ra nếu có mô hình trực giao trên thì:

$$\text{cov}(X) = \Sigma = LL^T + \psi \quad (5)$$

Như vậy việc tìm cấu trúc mô hình nhân tố trực giao đưa về việc tách ma trận hiệp phương sai dưới dạng (5).

Mô hình phân tích nhân tố trực giao

Nhận xét

Nếu L thỏa mãn (5) thì $L^* = LT$ cũng thỏa mãn với T là ma trận trực giao.

$$\Sigma = LL^T + \psi = LTT^TL^T + \psi = (LT)(T^TL^T) + \psi = (L^*)(L^*)^T + \psi$$

Như vậy hệ thức phân tích (5) xác định L sai khác một phép biến đổi trực giao.

Mô hình phân tích nhân tố trực giao

Từ (5) ta có:

$$\begin{aligned} D(X_i) &= l_{i1}^2 + \cdots + l_{im}^2 + \psi_i = \sigma_{ii} \\ \text{cov}(X_i, X_j) &= l_{i1}l_{j1} + \cdots + l_{im}l_{jm} = \sigma_{ij} \\ \text{cov}(X_i, F_j) &= l_{ij} \end{aligned} \tag{6}$$

$h_i^2 = l_{i1}^2 + \cdots + l_{im}^2$ gọi là phương sai chung,
 ψ_i được gọi là phương sai xác định.

Mô hình phân tích nhân tố trực giao

Ta có:

$$\underbrace{\sigma_{ii}}_{D(X_i)} = \underbrace{l_{i1}^2 + \cdots + l_{im}^2}_{\text{phương sai chung}} + \underbrace{\psi_i}_{\text{phương sai xác định}}$$

Mô hình phân tích nhân tố trực giao

Tóm lại, ứng với mô hình nhân tố trực giao, ta có cấu trúc sau đây về hiệp phương sai:

$$\textcircled{1} \text{ cov}(X) = \Sigma = LL^T + \psi$$

hoặc:

$$\begin{aligned} D(X_i) &= l_{i1}^2 + \cdots + l_{im}^2 + \psi_i \\ \text{cov}(X_i, X_j) &= l_{i1}l_{j1} + \cdots + l_{im}l_{jm} \end{aligned} \quad (7)$$

$$\textcircled{2} \text{ cov}(X, F) = L$$

hoặc:

$$\text{cov}(X_i, F_j) = l_{ij}$$

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng**
 - Phương pháp sử dụng phân tích thành phần chính
 - Phương pháp ước lượng hợp lý cực đại
 - Kiểm định mẫu lớn cho số lượng các nhân tố chung
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
 - Phương pháp sử dụng phân tích thành phần chính
 - Phương pháp ước lượng hợp lý cực đại
 - Kiểm định mẫu lớn cho số lượng các nhân tố chung
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng

Phương pháp sử dụng phân tích thành phần chính

Với phương pháp PCA, giả sử chúng ta có ma trận $\text{cov}(X) = \Sigma$ có p cặp trị riêng (Eigenvalue) $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ và hệ vector riêng (Eigenvector) trực chuẩn e_1, e_2, \dots, e_p là $(\lambda_1 e_1), \dots, (\lambda_p e_p)$. Khi đó, ta sẽ có:

$$\begin{aligned}\Sigma &= \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \dots + \lambda_p e_p e_p^T \\ &= \begin{bmatrix} \sqrt{\lambda_1} e_1 & \sqrt{\lambda_2} e_2 & \dots & \sqrt{\lambda_p} e_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} e_1^T \\ \sqrt{\lambda_2} e_2^T \\ \vdots \\ \sqrt{\lambda_p} e_p^T \end{bmatrix} \quad (8)\end{aligned}$$

Phương pháp sử dụng phân tích thành phần chính

Điều này phù hợp với cấu trúc hiệp phương sai quy định cho mô hình phân tích nhân tố có nhiều nhân tố là biến ($m = p$) và phương sai cụ thể $\psi_i = 0$ với mọi i . Ma trận tải có cột thứ j được biểu diễn bởi $\sqrt{\lambda_j} e_j$.

Như vậy, nếu thực hiện phân tích phổ của với $m = p$ nhân tố thì:

$$\sum_{(p \times p)} = \underset{(p \times p)}{L} \underset{(p \times p)}{L^T} + \underset{(p \times p)}{\psi} \quad (9)$$

→ Đây là mô hình p nhân tố cho $\underset{(p \times 1)}{X}$

Phương pháp sử dụng phân tích thành phần chính

Mặc dù đại diện cho phân tích nhân tố của ma trận là chính xác, tuy nhiên nó lại không thực sự hữu ích vì $m = p$. Việc sử dụng càng nhiều nhân tố chung càng có nhiều biến số trong khi đó không cho phép bất kỳ sự thay đổi nào trong các nhân tố cụ thể ϵ :

- Thay vào đó, người ta cố gắng đưa về mô hình nhân tố với $m < p$.
- Xét trường hợp với $m < p$ thì $p - m$ giá trị riêng cuối cùng của Σ là không đáng kể, tức là $\lambda_{m+1}, \dots, \lambda_p$ gần 0 hay trong phân tích (8) đại lượng $\lambda_1 e_1 e_1^\top + \dots + \lambda_m e_m e_m^\top$ có thể bỏ qua. Chúng ta có thể bỏ qua sự đóng góp của các giá trị riêng $\lambda_{m+1}, \dots, \lambda_p$ này cho Σ .

Phương pháp sử dụng phân tích thành phần chính

- Khi đó:

$$\begin{aligned}\sum &\approx \lambda_1 e_1 e_1^\top + \cdots + \lambda_m e_m e_m^\top = \begin{bmatrix} \sqrt{\lambda_1} e_1 & \cdots & \sqrt{\lambda_m} e_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} e_1^\top \\ \vdots \\ \sqrt{\lambda_m} e_m^\top \end{bmatrix} \\ &= \underset{(p \times m)}{L} \underset{(m \times p)}{L^\top}\end{aligned}$$

Phương pháp sử dụng phân tích thành phần chính

Phương sai của các nhân tố cụ thể có thể được coi là các phần tử trên đường chéo của ma trận $\sum - LL^T$.

Tức là ta có: $\psi_i = \sigma_{ii} - \sum_{j=1}^m l_{ij}^2, i = 1, \dots, p$

$$\Rightarrow \psi = \begin{pmatrix} \psi_1 & & 0 \\ & \ddots & \\ 0 & & \psi_p \end{pmatrix}$$

$$\Rightarrow \sum \approx LL^T + \psi \quad (10)$$

Phương pháp sử dụng phân tích thành phần chính

Ta có thể áp dụng phân tích (2.2) với dãy quan sát X_1, X_2, \dots, X_n và theo thông lệ dãy đó được quy tâm. Nói cách khác, thay vì xét dãy X_1, X_2, \dots, X_n ta xét:

- Tính vector trung bình mẫu quan sát được \bar{X} .
- Tính các vector độ lệch $X_j - \bar{X}$. Dãy X_1, X_2, \dots, X_n và dãy (2.3) có cùng ma trận hiệp phương sai mẫu S .
($p \times p$)
- Trong trường hợp các đơn vị của các biến không tương xứng, thường là mong muốn thực hiện với các biến được chuẩn hóa Z sau:

$$z_j = \begin{bmatrix} \frac{x_{j1} - \bar{x}_1}{\sqrt{s_{11}}} \\ \frac{x_{j2} - \bar{x}_2}{\sqrt{s_{22}}} \\ \vdots \\ \frac{x_{jp} - \bar{x}_p}{\sqrt{s_{pp}}} \end{bmatrix}, j = 1, \dots, n$$

Phương pháp sử dụng phân tích thành phần chính

Với Z có ma trận hiệp phương sai là R và đó chính là ma trận tương quan mẫu của các quan sát X_1, X_2, \dots, X_n . Tiêu chuẩn hóa tránh các vấn đề của việc có một biến với phương sai lớn ảnh hưởng quá mức đến việc tải nhân tố. Do các biến sau đều có kì vọng $\mu = 0$ và phương sai $\sigma^2 = 1$.

Sự biểu diễn (2.2) được gọi là PCA – Phương pháp phân tích thành phần chính khi chúng ta áp dụng cho ma trận hiệp phương sai mẫu S hoặc ma trận tương quan R .

Phân tích hai ma trận R và S:

- Phân tích nhân tố thành phần chính của ma trận hiệp phương sai mẫu được xác định theo các cặp trị riêng và vector riêng của S.
- Giả sử, chúng ta có các cặp trị riêng và vector riêng ấy $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2) \dots (\hat{\lambda}_p, \hat{e}_p)$ với $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$.
- Giả sử $\hat{\lambda}_{m+1}, \hat{\lambda}_{m+2} \geq \dots \hat{\lambda}_p$ là bé. Khi đó giả sử m là nhân tố ta muốn phân tích. Khi đó, ma trận tải trọng nhân tố $[l_{ij}]$ được ước lượng bởi:

$$[\sqrt{\lambda_1}e_1 \quad \dots \quad \sqrt{\lambda_m}e_m]$$

Phương pháp sử dụng phân tích thành phần chính

- Lúc đó, các phương sai xác định ψ_i được ước lượng bởi các phần tử trên đường chéo chính của ma trận $S - \hat{L}\hat{L}^\top$:

$$\hat{\psi}_i = \text{diag}(\psi_1, \dots, \psi_m) \text{ với } \psi_i = s_{ii} - \sum_{j=1}^m \hat{l}_{ij}^2$$

Phương pháp sử dụng phân tích thành phần chính

- Các phương sai chung được tính toán bởi:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2 = \sum_{j=1}^m \hat{l}_{ij}^2$$

⇒ Phân tích với ma trận tương quan mẫu R được thực hiện tương tự.

- Chúng ta xét ma trận dư $S - \hat{L}\hat{L}^\top$. Người ta luôn chứng minh được rằng tổng bình phương của các phần tử của $S - \hat{L}\hat{L}^\top$ bé hơn hoặc bằng $\hat{\lambda}_{m+1}^2 + \hat{\lambda}_{m+2}^2 \geq \dots \hat{\lambda}_p^2$.

Phương pháp sử dụng phân tích thành phần chính

Thuật toán

- Tính kỳ vọng của toàn bộ dữ liệu: $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$
- Trừ mỗi điểm dữ liệu đi vector kỳ vọng của toàn bộ dữ liệu:

$$\hat{x} = x_n - \bar{x}$$

- Tính ma trận hiệp phương sai: $S = \frac{1}{N} \hat{X} \cdot \hat{X}^T$
- Tính các trị riêng và vector riêng có norm bằng 1 của ma trận này, sắp xếp chúng theo thứ tự giảm dần của trị riêng.

Phương pháp sử dụng phân tích thành phần chính

Thuật toán

- Chọn K vector riêng ứng với K trị riêng lớn nhất để xây dựng ma trận U_k có các cột tạo thành một hệ trực giao, K vectors này, còn được gọi là các thành phần chính, tạo thành một không gian con gần với phân bố của dữ liệu ban đầu đã chuẩn hóa.
- Chiếu dữ liệu ban đầu đã chuẩn hóa \bar{X} xuống không gian con tìm được.
- Dữ liệu mới chính là tọa độ của các điểm dữ liệu trên không gian mới: $Z = U_k^T \cdot \hat{X}$ Dữ liệu ban đầu có thể tính được xấp xỉ theo dữ liệu mới như sau: $x \approx U_k Z + \bar{x}$

Phương pháp sử dụng phân tích thành phần chính

Lưu ý 1

Đối với giải pháp thành phần chính, ước lượng của tải trọng cho một nhân tố nhất định không thay đổi khi số lượng các nhân tố được tăng lên.

Tổng quát:

$$m = k : L_{(k)} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \dots & \sqrt{\hat{\lambda}_k} \hat{e}_k \end{bmatrix}$$

$$m = k + 1 : L_{(k+1)} = \begin{bmatrix} L_{(k)} & \vdots & \sqrt{\hat{\lambda}_{(k+1)}} e_{(k+1)} \end{bmatrix}$$

Phương pháp sử dụng phân tích thành phần chính

Ví dụ

$$m = 1 : L_{(1)} = \left[\sqrt{\hat{\lambda}_1} \hat{e}_1 \right]$$

$$m = 2 : L_{(2)} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 \end{bmatrix}$$

Phương pháp sử dụng phân tích thành phần chính

Lưu ý 2

Sự đóng góp của các nhân tố vào tổng phương sai mẫu.

Phương pháp sử dụng phân tích thành phần chính

- Ta có:

$$s_{ii} = \sum_{j=1}^m \hat{l}_{ij}^2 + \hat{\psi}_i$$

- 1 Sự đóng góp của nhân tố đầu tiên cho s_{ii} là \hat{l}_{i1}^2 .
- 2 Sự đóng góp của nhân tố đầu tiên cho tổng phương sai mẫu $tr(S) = s_{11} + s_{22} + \dots + s_{33}$ là:

$$\hat{l}_{11} + \hat{l}_{22} + \dots + \hat{l}_{p1} \quad (11)$$

Phương pháp sử dụng phân tích thành phần chính

- Với: $\hat{L}_{(p \times m)} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix}$

thì ta có cột thứ j : $\begin{bmatrix} \hat{l}_{1j} \\ \vdots \\ \hat{l}_{pj} \end{bmatrix} = \sqrt{\hat{\lambda}_j} \hat{e}_j$

- Chẳng hạn $j = 1$ thì ta có cột đầu tiên của \hat{L} : $\begin{bmatrix} \hat{l}_{11} \\ \vdots \\ \hat{l}_{p1} \end{bmatrix}$

- Như vậy:

$$\sum_{i=1}^p \hat{l}_{ij}^2 = (\sqrt{\hat{\lambda}_j} \hat{e}_j)^\top (\sqrt{\hat{\lambda}_j} \hat{e}_j) = \hat{\lambda}_j \hat{e}_j^\top \hat{e}_j = \hat{\lambda}_j$$

Phương pháp sử dụng phân tích thành phần chính

Ví dụ

$$j = 1 \Rightarrow \sum_{i=1}^p \hat{l}_{i1}^2 = \hat{\lambda}_1$$

- Khi đó tỷ lệ tổng phương sai mẫu được giải thích thông qua nhân tố đầu tiên:

$$\frac{\hat{\lambda}_1}{tr(S)} = \frac{\hat{\lambda}_1}{\left(\sum_{i=1}^p s_{ii}\right)}$$

Phương pháp sử dụng phân tích thành phần chính

Ví dụ

- Tỷ lệ tổng phương sai mẫu được giải thích thông qua k nhân tố đầu tiên:

$$\frac{\sum_{i=1}^k \hat{\lambda}_i}{\sum_{i=1}^p s_{ii}}$$

Phương pháp sử dụng phân tích thành phần chính

Ví dụ 3.1: Phân tích nhân tố cho ưu tiên của người tiêu dùng

Trong một nghiên cứu ưu tiên của người tiêu dùng, trong một mẫu ngẫu nhiên của khách hàng đã được yêu cầu đánh giá một số thuộc tính của một sản phẩm mới. Các phản hồi, trên thang phân biệt chữ nghĩa 7 điểm, đã được lập bảng và ma trận R tương quan các thuộc tính được xây dựng.

Thuộc tính	Mùi vị	Giá thành	Hương vị	Ăn nhanh	Năng lượng
Mùi vị	1.000	0.02	0.96	0.42	0.01
Giá thành	0.02	1.000	0.13	0.71	0.85
Hương vị	0.96	0.13	1.000	0.50	0.11
Ăn nhanh	0.42	0.71	0.50	1.000	0.79
Năng lượng	0.01	0.85	0.11	0.79	1.000

Phương pháp sử dụng phân tích thành phần chính

Nhận xét: Chúng ta có thể thấy rằng các biến mùi vị và hương vị có hệ số tương quan khá cao, tương tự với biến giá thành và năng lượng. Từ trên, ta có mong muốn rằng quan hệ giữa các biến có thể được giải thích qua hai hoặc ba biến.

- Tính toán các giá trị riêng của ma trận R, ta thấy hai thành phần đầu tiên đó: $\lambda_1 = 2.85$ và $\lambda_2 = 1.81$ là hai giá trị riêng duy nhất lớn hơn 1. Khi đó:

$$\frac{\lambda_1 + \lambda_2}{\sum_{i=1}^5 \lambda_i} = 0.93$$

Phương pháp sử dụng phân tích thành phần chính

- Như vậy, tổng $\lambda_1 + \lambda_2$ chiếm 93% tổng phương sai của 5 biến đã chuẩn hoá. Ta có ma trận hệ số tải L:

$$L = \begin{bmatrix} & \sqrt{\lambda_1} e_1 & \sqrt{\lambda_2} e_2 \\ \text{Mùi vị} & 0.56 & 0.82 \\ \text{Giá thành} & 0.78 & -0.53 \\ \text{Hương liệu} & 0.65 & 0.75 \\ \text{Ăn nhanh} & 0.94 & -0.10 \\ \text{Năng lượng} & 0.80 & -0.54 \end{bmatrix}$$

- Tuy nhiên các hệ số tải khá là cao. Ví dụ ở hương liệu có vẻ quan trọng với cả yếu tố 1 và 2. Điều này không cung cấp một cách giải thích dữ liệu đơn giản và rõ ràng. Lý tưởng nhất là mỗi biến sẽ xuất hiện như một yếu tố đóng góp đáng kể trong một cột.

Phương pháp sử dụng phân tích thành phần chính

Ví dụ 3.2: Phân tích nhân tố dữ liệu giá cổ phiếu

Tỷ suất lợi nhuận hàng tuần của 5 cổ phiếu (JPMorgan, Citibank, Wells Fargo, Royal Dutch Shell và ExxonMobil) niêm yết trên Sở giao dịch chứng khoán New York được xác định trong khoảng thời gian từ tháng 1/2004 đến tháng 12/2005. Tỷ suất lợi nhuận hàng tuần được xác định là:

$$\frac{\text{Giá đóng cửa tuần hiện tại} - \text{Giá đóng cửa tuần trước}}{\text{Giá đóng cửa tuần trước}}$$

được điều chỉnh theo việc chia cổ phiếu và cổ tức. Các quan sát trong 103 tuần liên tiếp dường như được phân bố độc lập, nhưng tỷ suất lợi nhuận giữa các cổ phiếu có mối tương quan với nhau, bởi vì như người ta mong đợi, các cổ phiếu có xu hướng biến động cùng nhau để đáp ứng với các điều kiện kinh tế chung.

Phương pháp sử dụng phân tích thành phần chính

- Gọi x_1, x_2, \dots, x_5 lần lượt biểu thị tỷ suất lợi nhuận quan sát được hàng tuần của JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell và ExxonMobil.
- Ta có ma trận tương quan R :

$$R = \begin{bmatrix} 1.000 & 0.632 & 0.511 & 0.115 & 0.155 \\ 0.632 & 1.000 & 0.574 & 0.322 & 0.213 \\ 0.511 & 0.574 & 1.000 & 0.183 & 0.146 \\ 0.115 & 0.322 & 0.183 & 1.000 & 0.683 \\ 0.155 & 0.213 & 0.146 & 0.683 & 1.000 \end{bmatrix}$$

Phương pháp sử dụng phân tích thành phần chính

- Tính toán các giá trị riêng của ma trận R, ta thấy hai thành phần đầu tiên $\hat{\lambda}_1=2.437$ và $\hat{\lambda}_2=1.407$ là hai giá trị riêng duy nhất lớn hơn.
- Khi đó:

$$\left(\frac{\lambda_1 + \lambda_2}{p}\right) \cdot 100\% = \left(\frac{2.437 + 1.407}{5}\right) \cdot 100\% = 77\%$$

Phương pháp sử dụng phân tích thành phần chính

- Hệ số ma trận tải L:

$$L = \begin{bmatrix} & \sqrt{\lambda_1} e_1 & \sqrt{\lambda_2} e_2 \\ JP\ Morgan & 0.732 & -0.437 \\ Citibank & 0.831 & -0.280 \\ Wells\ Fargo & 0.726 & -0.374 \\ Royal\ Dutch\ Shell & 0.605 & 0.694 \\ ExxonMobil & 0.563 & 0.719 \end{bmatrix}$$

Phương pháp sử dụng phân tích thành phần chính

Nhận xét

- Yếu tố đầu tiên, đại diện cho các điều kiện kinh tế chung và có thể được gọi là yếu tố thị trường. Tất cả các cổ phiếu đều có mức độ ảnh hưởng lớn đến yếu tố này và mức độ này là như nhau.
- Yếu tố thứ hai so sánh cổ phiếu ngân hàng với cổ phiếu dầu mỏ. Do đó, nó phân biệt các cổ phiếu trong các ngành khác nhau và có thể được gọi là yếu tố ngành.
- Tóm lại, tỷ suất lợi nhuận dường như được xác định bởi các điều kiện và hoạt động thị trường chung dành riêng cho các ngành khác nhau, cũng như yếu tố dư thừa hoặc yếu tố cụ thể của công ty.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
 - Phương pháp sử dụng phân tích thành phần chính
 - Phương pháp ước lượng hợp lý cực đại
 - Kiểm định mẫu lớn cho số lượng các nhân tố chung
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng

Phương pháp ước lượng hợp lý cực đại

- Ước lượng hợp lý cực đại (MLE - Maximum Likelihood Estimation) là một phương pháp trong thống kê dùng để ước lượng giá trị tham số của một mô hình xác suất dựa trên những dữ liệu quan sát được.
- Phương pháp này ước lượng các tham số nói trên bởi những giá trị làm cực đại hóa hàm hợp lý.

Phương pháp ước lượng hợp lý cực đại

Nếu các nhân tố chung F và các nhân tố cụ thể ϵ có thể được coi là có phân phối chuẩn, khi đó ước lượng hợp lý cực đại của nhân tố tài có thể đạt được.

Khi F_j và ϵ_j đều có phân phối chuẩn, các quan sát $X_j - \mu = LF_j + \epsilon_j$ cũng là phân phối chuẩn, khi đó ta có hàm hợp lý cực đại là:

$$\begin{aligned} L(\mu, \Sigma) &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right) \text{tr} \left[\Sigma^{-1} \left(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^\top + n(\bar{\mathbf{x}} - \mu)(\bar{\mathbf{x}} - \mu)^\top \right) \right]} \\ &= (2\pi)^{-\frac{(n-1)p}{2}} |\Sigma|^{-\frac{(n-1)}{2}} e^{-\left(\frac{1}{2}\right) \text{tr} \left[\Sigma^{-1} \left(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^\top \right) \right]} \\ &\quad \times (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\left(\frac{n}{2}\right) (\bar{\mathbf{x}} - \mu)^\top \Sigma^{-1} (\bar{\mathbf{x}} - \mu)} \end{aligned} \quad (12)$$

phụ thuộc vào L và Ψ do $\Sigma = LL^\top + \Psi$.

Phương pháp ước lượng hợp lý cực đại

Mô hình này chưa được xác định cụ thể, do có nhiều phương án cho L có thể thực hiện được bằng các phép biến đổi trực giao. Điều mong muốn là làm cho L được xác định rõ ràng bằng cách áp đặt điều kiện thuận lợi, một trong số đó là điều kiện duy nhất.

$$L^T \Psi^{-1} L = \Delta \text{ là ma trận đường chéo} \quad (13)$$

Ước lượng hợp lý cực đại L và Ψ sẽ có được khi ta cực đại hóa hàm hợp lý bên trên. Chúng ta sẽ có những chương trình máy tính có sẵn để giải. Ta sẽ tổng hợp một số kết quả về ước lượng hợp lý cực đại.

Phương pháp ước lượng hợp lý cực đại

Cho X_1, X_2, \dots, X_n là mẫu ngẫu nhiên từ $N_p(\mu, \Sigma)$, ở đây $\Sigma = LL^\top + \Psi$ là ma trận hiệp phương sai của mô hình m nhân tố chung. Ước lượng hợp lý cực đại L, Ψ và $\mu = \bar{x}$ với điều kiện:

$$L^\top \Psi^{-1} L = \Delta \text{ là ma trận đường chéo}$$

Phương pháp ước lượng hợp lý cực đại

Ước lượng hợp lý cực đại cho tính cộng đồng là:

$$h_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2 \text{ với } i = 1, 2, \dots, p \quad (14)$$

hay

$$\left(\begin{array}{c} \text{Tỷ lệ đóng góp của nhân tố } j \\ \text{trong tổng phương sai} \end{array} \right) = \frac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \dots + \hat{l}_{pj}^2}{s_{11} + s_{22} + \dots + s_{pp}} \quad (15)$$

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng**
 - Phương pháp sử dụng phân tích thành phần chính
 - Phương pháp ước lượng hợp lý cực đại
 - Kiểm định mẫu lớn cho số lượng các nhân tố chung
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng

Kiểm định mẫu lớn cho số lượng các nhân tố chung

Giả sử ta có mô hình m nhân tố chung. Trong trường hợp $\Sigma = LL^T + \Psi$, kiểm định tính đầy đủ của mô hình m nhân tố chung tương đương với kiểm định giả thuyết:

$$H_0 : \underset{(p \times p)}{\Sigma} = \underset{(p \times m)}{L} \underset{(m \times p)}{L^T} + \underset{(p \times p)}{\Psi}$$

và đối thuyết $H_1 : \Sigma$ là bất kì ma trận xác định dương nào khác.

Kiểm định mẫu lớn cho số lượng các nhân tố chung

Khi Σ không có các điều kiện đặc biệt, giá trị lớn nhất của hàm hợp lý (với $\hat{\Sigma} = \frac{n-1}{n}S = S_n$) tỷ lệ thuận với:

$$|S_n|^{-\frac{n}{2}} e^{-\frac{np}{2}}$$

Kiểm định mẫu lớn cho số lượng các nhân tố chung

Theo $H_0, \Sigma = LL^\top + \Psi$, trong trường hợp này, thay $\hat{\mu} = \bar{x}$ vào công thức hàm hợp lý cực đại tổng quát, ở đây \hat{L} và $\hat{\Psi}$ lần lượt là ước lượng hợp lý cực đại của L và Ψ , ta có giá trị lớn nhất của hàm hợp lý tỷ lệ thuận với:

$$\begin{aligned} & |\hat{\Sigma}|^{-n/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\hat{\Sigma}^{-1} \left(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (\mathbf{x}_j - \bar{\mathbf{x}})^\top \right) \right] \right\} \\ &= |\hat{\mathbf{L}}\hat{\mathbf{L}}^\top + \hat{\Psi}|^{-n/2} \exp \left\{ -\frac{1}{2} n \text{tr} [(\hat{\mathbf{L}}\hat{\mathbf{L}} + \hat{\Psi})^{-1} \mathbf{S}_n] \right\} \end{aligned}$$

Kiểm định mẫu lớn cho số lượng các nhân tố chung

Kết hợp hai điều trên, ta tìm được thống kê tỷ lệ khả năng cho kiểm tra H_0 là:

$$\begin{aligned} -2 \ln \Lambda &= -2 \ln \left[\frac{\text{cực đại hợp lý ở } H_0}{\text{Cực đại hợp lý}} \right] \\ &= -2 \ln \left(\frac{|\hat{\Sigma}|}{|\mathbf{S}_n|} \right)^{-n/2} + n \left[\text{tr} \left(\hat{\Sigma}^{-1} \mathbf{S}_n \right) - p \right] \end{aligned}$$

với bậc tự do:

$$\begin{aligned} v - v_0 &= \frac{1}{2} p(p+1) - [p(m+1) - \frac{1}{2} m(m-1)] \\ &= \frac{1}{2} [(p-m)^2 - p - m] \end{aligned}$$

Kiểm định mẫu lớn cho số lượng các nhân tố chung

Do $\text{tr}(\hat{\Sigma}^{-1}S_n) - p = 0$ nên $\hat{\Sigma} = \hat{L}\hat{L}^\top + \hat{\Psi}$ là ước lượng hợp lý cực đại của $\Sigma = LL^\top + \Psi$. Như vậy, ta có:

$$-2 \ln \Lambda = n \ln \frac{|\hat{\Sigma}|}{|S_n|}$$

Kiểm định mẫu lớn cho số lượng các nhân tố chung

Bartlett đã chỉ ra rằng xấp xỉ chi bình phương đối với phân phối lấy mẫu của $-2\ln(\Lambda)$ trong công thức trên có thể được cải thiện bằng cách thay thế n bằng hệ số nhân $\left(n - 1 - \frac{2p+4m+5}{6}\right)$. Sử dụng hiệu chỉnh Bartlett, ta bác bỏ H_0 ở mức ý nghĩa α nếu:

$$\left(n - 1 - \frac{2p + 4m + 5}{6}\right) \ln \frac{|\hat{L}\hat{L}^\top + \hat{\Psi}|}{|S_n|} > \chi^2_{\frac{(p-m)^2 - p - m}{2}}(\alpha) \quad (16)$$

với n và $n - p$ lớn. Do số bậc tự do $\frac{(p-m)^2 - p - m}{2} > 0$ nên:

$$m < \frac{1}{2} \left(2p + 1 - \sqrt{8p + 1}\right)$$

Kiểm định mẫu lớn cho số lượng các nhân tố chung

Kết luận

- Kiểm định độ cầu của Bartlett (Bartlett's test of sphericity) dùng để xem xét các biến quan sát trong nhân tố có tương quan với nhau hay không bằng cách sử dụng ma trận tương quan quan sát với ma trận đơn vị.
- Điều kiện cần để áp dụng phân tích nhân tố là các biến quan sát phản ánh những khía cạnh khác nhau của cùng một nhân tố phải có mối tương quan với nhau.
- Chương trình tính ra xấp xỉ chi bình phương và giá trị $\text{sig}(P)$. Nếu $\text{sig} < 0.05$ ta nói các biến có tương quan với nhau và bộ dữ liệu có ý nghĩa thống kê và có thể phân tích nhân tố.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố**
 - Phép quay nhân tố
 - Phương pháp quay nhân tố trực giao
 - Các phương pháp quay nhân tố phân tích
 - Phương pháp quay xiên
- 5 Điểm nhân tố
- 6 Vận dụng

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
 - Phép quay nhân tố
 - Phương pháp quay nhân tố trực giao
 - Các phương pháp quay nhân tố phân tích
 - Phương pháp quay xiên
- 5 Điểm nhân tố
- 6 Vận dụng

Phép quay nhân tố

Như chúng ta đã chỉ ra trong (2) (Mô hình nhân tố trực giao), tất cả các tải nhân tố thu được từ các tải ban đầu bằng một **phép biến đổi trực giao** đều có khả năng tái tạo lại ma trận hiệp phương sai (hoặc ma trận tương quan) như nhau (3).

Từ đại số ma trận, chúng ta biết rằng một phép biến đổi trực giao thì tương ứng với một phép quay cứng nhắc tọa độ của các trục.

Phép quay nhân tố

Định nghĩa 4.1

Phép quay nhân tố là phép biến đổi tạo ra ma trận tải nhân tố quay vòng L^* để có một cách giải thích dữ liệu đơn giản và dễ dàng hơn.

Phép quay nhân tố có 2 loại:

- Phép quay trực giao: Trong đó các nhân tố chung không tương quan với nhau.
- Phép quay xiên: Trong đó các nhân tố chung có tương quan với nhau.

⇒ Phép quay xiên thích hợp hơn phép quay trực giao, bởi nó có xu hướng cung cấp các mẫu tải nhân tố dễ hiểu hơn mà không có các hạn chế phi thực tế rằng các nhân tố phổ biến không tương quan với nhau.

Định nghĩa 4.2

Ma trận của các tải quay nếu L là ma trận $p \times m$ của hệ số tải ước tính thu được bằng bất kì phương pháp nào (thành phần chính hay ước lượng hợp lý cực đại) thì:

$$L^* = LT \text{ trong đó } TT^T = T^T T = 1 \quad (17)$$

L^* là một ma trận $p \times m$ của các tải "quay".

Phép quay nhân tố

Hơn nữa, ma trận hiệp phương sai (hoặc tương quan) ước tính vẫn không thay đổi vì:

$$\hat{L}\hat{L}^\top + \hat{\Psi} = \hat{L}TT^\top\hat{L} + \hat{\Psi} = \hat{L}^*\hat{L}^{*T} + \hat{\Psi} \quad (18)$$

Phương trình (18) chỉ ra rằng ma trận dư:

$$S_n - \hat{L}\hat{L}^\top - \hat{\Psi} = S_n - \hat{L}^*\hat{L}^{*T} - \hat{\Psi}$$

không đổi sau khi thực hiện quay nhân tố. Hơn nữa các phương sai cụ thể $\hat{\Psi}_i$, và cộng đồng \hat{h}_i không thay đổi.

Kết luận

Do đó từ quan điểm toán học việc thu được \hat{L} hay \hat{L}^\top là không quan trọng. Vì tải trọng ban đầu có thể không dễ hiểu, nên thông thường, người ta sẽ quay chúng cho đến khi đạt được "cấu trúc đơn giản hơn".

\Rightarrow Lý tưởng nhất là chúng ta muốn thấy một mô hình tải trọng sao cho mỗi biên chịu tải trọng cao đối với một nhân tố duy nhất và có tải trọng nhỏ đến trung bình đối với các nhân tố còn lại.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố**
 - Phép quay nhân tố
 - Phương pháp quay nhân tố trực giao**
 - Các phương pháp quay nhân tố phân tích
 - Phương pháp quay xiên
- 5 Điểm nhân tố
- 6 Vận dụng

Phương pháp quay nhân tố trực giao

Tập trung vào các phương pháp đồ thị và phân tích để xác định vào phép quay trực giao cho một cấu trúc đơn giản.

Các yếu tố không tương quan được coi là các vector đơn vị dọc theo các trục tọa độ vuông góc. Biểu đồ của các cặp hệ số tải (C) tạo thành p điểm mỗi điểm tương ứng với một biến, sau đó tọa độ có thể được quay một cách trực giao thông qua một góc - gọi nó là Φ và tải trọng mới l_{ij}^* được xác định từ các mối quan hệ:

$$\hat{L}_{(p \times 2)}^* = \hat{L}_{(p \times 2)} T_{(2 \times 2)} \quad (19)$$

Phương pháp quay nhân tố trực giao

Trong đó:

$$\left\{ \begin{array}{l} T = \begin{bmatrix} \cos \Phi & \sin \Phi \\ -\sin \Phi & \cos \Phi \end{bmatrix} \\ T = \begin{bmatrix} \cos \Phi & -\sin \Phi \\ \sin \Phi & \cos \Phi \end{bmatrix} \end{array} \right. \begin{array}{l} \text{quay theo chiều kim đồng hồ} \\ \text{quay ngược chiều kim đồng hồ} \end{array}$$

- Khi $m = 2$ hoặc các thừa số chung được coi là hai nhân tử cùng một lúc, sự biến đổi thành một cấu trúc đơn giản thường có thể được xác định bằng đồ thị.
- Đối với $m > 2$, các định hướng không dễ dàng hình dung được và độ lớn của các tải trọng quay phải được kiểm tra để tìm ra cách giải thích có ý nghĩa đối với dữ liệu gốc.

Phương pháp quay nhân tố trực giao

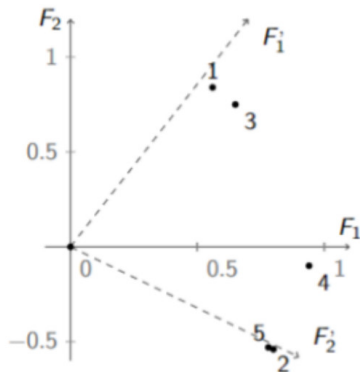
Ví dụ 4.1

Chúng ta xét lại ví dụ (4.1) để hiểu thêm về phương pháp quay nhân tố:

$$\mathbf{L} = \begin{bmatrix} & \lambda_1 \sqrt{e_1} & \lambda_1 \sqrt{e_1} \\ \text{Mùi vị} & 0.56 & 0.82 \\ \text{Giá thành} & 0.78 & -0.53 \\ \text{Hương liệu} & 0.65 & 0.75 \\ \text{Ăn nhanh} & 0.94 & -0.10 \\ \text{Nướng lượng} & 0.80 & -0.54 \end{bmatrix}$$

⇒ Chúng ta sẽ tìm cách quay nhân tố sao cho mỗi biến sẽ xuất hiện như một yếu tố đóng góp đáng kể trong một cột. Để có thể giải thích dữ liệu một cách đơn giản và rõ ràng hơn.

Phương pháp quay nhân tố trực giao



Hình: Đồ thị Biểu diễn ma trận tải L

Phương pháp quay nhân tố trực giao

Các cặp tải nhân tố được vẽ trên hệ tọa độ OF_1F_2 , hệ tọa độ OF_1F_2 là hệ tọa độ sau khi quay:

$$L' = \begin{bmatrix} \text{Hương vị} & 0.02 & 0.99 \\ \text{Giá} & 0.87 & 0.0065 \\ \text{Mùi hương} & 0.13 & 0.96 \\ \text{Nhanh} & 0.81 & 0.4 \\ \text{Năng lượng} & 0.97 & -0.016 \end{bmatrix}$$

Yếu tố 2 có tải trọng cao với vị và hương có thể là yếu tố về hương vị, yếu tố 1 có tải cao với giá, ăn nhanh, năng lượng có thể là yếu tố quyết định về giá thành và dinh dưỡng.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
 - Phép quay nhân tố
 - Phương pháp quay nhân tố trực giao
 - Các phương pháp quay nhân tố phân tích
 - Phương pháp quay Quartimax
 - Phương pháp quay Varimax
 - Phương pháp quay xiên
- 5 Điểm nhân tố
- 6 Vận dụng

Các phương pháp quay nhân tố phân tích

Phương pháp quay nhân tố khách quan đầu tiên được đưa ra bởi Carroll (1953), mặc dù một số phương pháp đã được đưa ra ngay sau đó, trong trường hợp trực giao, được cho là tương đương với giải pháp của Carroll. Thuật ngữ chung cho các giải pháp tương đương này là "Phương pháp Quartimax".

Thủ tục Varimax của Kaiser (1958) là một sửa đổi của thủ tục Quartimax và có lẽ là phương pháp phân tích quay nhân tố được sử dụng phổ biến nhất.

Quartimax và Varimax là những trường hợp đặc biệt của lớp tiêu chuẩn trực giao cho phép quay trực giao.

Các phương pháp quay xiên đầu tiên được gọi là các phương pháp gián tiếp vì chúng liên quan đến sử dụng các cấu trúc tham chiếu.

Định nghĩa 4.3

Một phương pháp quay không đơn lẻ không làm thay đổi lượng phương sai được giải thích và tính cộng đồng của từng biến cũng không thay đổi.

Nghĩa là:

$$\left(\sum_{j=1}^q \lambda_{ij}^2 \right)^2 = \sum_{j=1}^q \lambda_{ij}^4 + \sum_{j=1}^q \sum_{j \neq k}^q \lambda_{ij}^2 \lambda_{ik}^2 = \text{constant}, \quad (20)$$

chỉ số $j = 00$ đã được bỏ qua để biểu thức được rõ ràng hơn.

Phương pháp quay Quartimax

Tổng hợp phương trình (20) trên tất cả các biến p đã cho:

$$\sum_{i=1}^p \sum_j = 1^q \lambda_{ij}^4 + \sum_{i=1}^p \sum_{j=1}^q \sum_{j \neq k}^q \lambda_{ij}^2 \lambda_{ik}^2 = \text{constant} \quad (21)$$

Một ví dụ về phân tích nhân tố được đưa ra dưới đây chỉ ra rằng phương pháp Quartimax có xu hướng giữ lại một phần nhân tố đầu tiên quan trọng. Đây là đặc điểm của phương pháp luân chuyển nhân tố và xảy ra bởi vì quartimax về cơ bản cố gắng đơn giản hóa các hàng của ma trận mẫu thông qua việc giảm thiểu số hạng tích chéo trong phương trình (21).

Phương pháp quay Varimax

Phương pháp Varimax cố gắng đơn giản hóa các cột thay vì các hàng của ma trận mẫu. Do đó, nó ngăn cản việc duy trì một yếu tố đầu tiên khá chung chung.

Kaiser (1958) định nghĩa tính đơn giản của một nhân tố là bình phương của phương sai của tải trọng của nó. Đối với yếu tố này người ta biểu diễn bằng biến v_j^* , trong đó:

$$v_j^* = \frac{1}{p} \left[\sum_{i=1}^p \lambda_{ij}^4 - \frac{1}{p} \left(\sum_{i=1}^p \lambda_{ij}^2 \right)^2 \right], j = 1, \dots, q. \quad (22)$$

Phương pháp quay Varimax

Phương pháp Varimax liên quan đến việc tối đa hóa tổng số đơn giản, tức là tối đa hóa:

$$V^* = \sum_{j=1}^q v_j^* \quad (23)$$

Phương pháp quay Varimax

Trong thực tế, các hệ số tải nhân tố thường được chuẩn hóa bởi các cộng đồng tương ứng của chúng, không thay đổi bởi một vòng quay không đơn lẻ. Do đó tiêu chí Varimax được đưa ra bởi:

$$V = \frac{1}{p} \sum_{j=1}^q \left[\sum_{i=1}^p \frac{\lambda_{ij}^4}{h_i^4} - \frac{1}{p} \left(\sum_{i=1}^p \frac{\lambda_{ij}^2}{h_i^2} \right)^2 \right], \quad (24)$$

trong đó:

$$h_i^2 = \sum_{j=1}^q \lambda_{ij}^2, i = 1, \dots, p, \quad (25)$$

là tính cộng đồng của biến x . Như với phép quay quartimax, quy trình tính toán liên quan đến phép quay theo cặp của các thừa số.

Kiểm tra tỷ lệ phương sai tương đối được giải thích bởi từng yếu tố chỉ ra rằng Varimax có xu hướng làm các tải lớn trên các cột của ma trận mẫu ở mức độ lớn hơn so với quartimax. Điều này là do, để V đạt cực đại, số hạng thứ hai trong phương trình (25) phải nhỏ, tức là:

$$\sum_{i=1}^p \frac{\lambda_{ij}^2}{h_i^2}$$

phải tương đối ổn định giữa các yếu tố.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố**
 - Phép quay nhân tố
 - Phương pháp quay nhân tố trực giao
 - Các phương pháp quay nhân tố phân tích
 - **Phương pháp quay xiên**
- 5 Điểm nhân tố
- 6 Vận dụng

Định nghĩa 4.4

"Oblimin trực tiếp" là thủ tục quay xiên được giới thiệu bởi Jennrich và Sampson (1966) không liên quan đến việc sử dụng các trục tham chiếu. Tiêu chí oblimin trực tiếp có dạng như sau:

$$\sum \sum_{j < k}^q \left[\sum_{i=1}^p \lambda_{ij}^2 \lambda_{ik}^2 - \frac{\zeta}{p} \sum_{i=1}^p \lambda_{ij}^2 \sum_{i=1}^p \lambda_{ik}^2 \right], \quad (26)$$

trong đó các chỉ số 0 đã được bỏ qua cho rõ ràng. ζ là một tham số kiểm soát mức độ tương quan giữa các yếu tố.

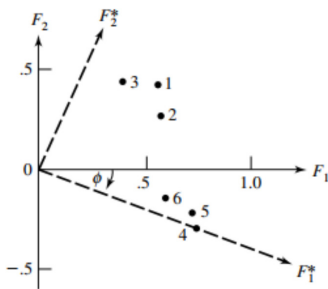
Phương pháp quay xiên

- Các phép quay trực giao phù hợp với mô hình nhân tố trong đó các nhân tố chung được giả định là độc lập.
- Nhiều nhà nghiên cứu trong khoa học xã hội xem xét phép quay xiên (không trực giao).
- Nếu coi m thừa số chung là các trục tọa độ thì điểm có m tọa độ $(\hat{l}_{i1}, \hat{l}_{i2}, \dots, \hat{l}_{im})$ biểu thị vị trí của biến thứ i trong không gian nhân tử.

Phương pháp quay xiên

- Giả sử rằng các biến được nhóm thành các cụm không chồng chéo, một phép quay trực giao đến một cấu trúc đơn giản tương ứng với một phép quay cứng nhắc của các trục tọa độ sao cho các trục, sau khi quay, đi càng gần các cụm càng tốt.
- Một phép quay xiên đối với một cấu trúc đơn giản tương ứng với một phép quay không cứng nhắc của hệ tọa độ sao cho các trục quay (không còn vuông góc) đi qua (gần như) qua các cụm.
- Một phép quay xiên tìm cách thể hiện từng biến theo thuật ngữ của một số yếu tố tối thiểu tốt nhất là một yếu tố duy nhất.

Phương pháp quay xiên



Hình: Phép quay xiên

Nhận xét: Trong thực tế người ta sử dụng phương pháp quay xiên phổ biến hơn vì nó sẽ cho kết quả một cách trực quan và dễ hiểu hơn và các nhân tố chung tương quan với nhau.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố**
 - Phương pháp bình phương tối thiểu có trọng số
 - Phương pháp hồi quy
- 6 Vận dụng

Điểm nhân tố

Trong phân tích nhân tố, ta thường quan tâm vào các tham số trong mô hình nhân tố. Tuy nhiên, các giá trị ước lượng của các nhân tố chung, gọi là điểm nhân tố, cũng có thể được yêu cầu. Những đại lượng này thường được sử dụng cho mục đích chuẩn đoán, cũng như đầu vào cho phân tích tiếp theo.

Điểm nhân tố không phải là ước lượng của các tham số chưa biết theo nghĩa thông thường. Thay vào đó, chúng là ước lượng của các giá trị cho các vector nhân tố ngẫu nhiên không quan sát được. Tức là điểm nhân tố:

$\hat{f}_j =$ ước lượng của giá trị f_j đạt được bởi F_j (thành phần thứ j).

Việc ước lượng này rất phức tạp bởi vì số lượng các đại lượng không quan sát được f_j và ϵ_j nhiều hơn số đại lượng quan sát được x_i . Để vượt qua trở ngại này, ta có một số cách giải quyết. Chương này sẽ đề cập đến 2 cách tiếp cận chủ yếu là phương pháp bình phương tối thiểu có trọng số và phương pháp hồi quy.

Hai cách tiếp cận đều có 2 yếu tố chung:

- Ta coi ước lượng của các hệ số tải \hat{l}_{ij} và phương sai cụ thể $\hat{\Psi}_j$ là các giá trị thực.
- Chúng liên quan đến các phép biến đổi tuyến tính của dữ liệu gốc, có thể đã được chuẩn hóa.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố**
 - Phương pháp bình phương tối thiểu có trọng số
 - Phương pháp hồi quy
- 6 Vận dụng

Phương pháp bình phương tối thiểu có trọng số

Trước hết, ta giả sử rằng vector trung bình μ , ma trận tải L và vector phương sai cụ thể Ψ đã biết với mô hình nhân tố:

$$\underset{(p \times 1)}{X} - \underset{(p \times 1)}{\mu} = \underset{(p \times m)}{L} \underset{(m \times 1)}{F} + \underset{(p \times 1)}{\epsilon}$$

Hơn nữa, ta coi các sai số $\epsilon^T = [\epsilon_1, \epsilon_2, \dots, \epsilon_p]$ là các nhân tố xác định. Vì $Var(\epsilon_i) = \Psi_i, i = 1, 2, \dots, p$, không cần phải bằng nhau, Bartlett đã gợi ý rằng phương pháp bình phương tối thiểu có trọng số có thể được sử dụng để ước lượng giá trị các nhân tố chung.

Phương pháp bình phương tối thiểu có trọng số

Tổng bình phương của các nhân tố xác định, có trọng số bằng nghịch đảo các phương sai của chúng là:

$$\sum_{i=1}^P \frac{\epsilon_i^2}{\psi_i} = \epsilon^\top \Psi^{-1} \epsilon = (x - \mu - Lf)^\top \Psi^{-1} (x - \mu - Lf) \quad (27)$$

Bartlett đề xuất chọn các ước lượng của f để cực tiểu hóa (27). Ta thu được nghiệm là:

$$\hat{f} = (L^\top \Psi^{-1} L)^{-1} L^\top \Psi^{-1} (x - \mu) \quad (28)$$

Phương pháp bình phương tối thiểu có trọng số

Từ công thức (28), ta lấy các ước lượng \hat{L} , $\hat{\Psi}$, và $\hat{\mu} = \bar{x}$ là các giá trị thực và thu được điểm nhân tố cho thành phần thứ j là:

$$\hat{f}_j = (\hat{L}^\top \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}^\top \hat{\Psi}^{-1} (x_j - \bar{x}) \quad (29)$$

Khi \hat{L} và $\hat{\Psi}$ được xác định bằng phương pháp ước lượng hợp lí cực đại thì nó phải thỏa mãn điều kiện duy nhất, $\hat{L}^{-1\top} \hat{\Psi}^{-1} \hat{L} = \hat{\Delta}$ là một ma trận đường chéo. Từ đó ta có kết quả sau đây:

Phương pháp bình phương tối thiểu có trọng số

Điểm nhân tố thu được bằng phương pháp bình phương tối thiểu có trọng số từ ước lượng hợp lý cực đại:

$$\begin{aligned}\hat{f}_j &= (\hat{L}^\top \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}^\top \hat{\Psi}^{-1} (x_j - \hat{\mu}) \\ &= \hat{\Delta}^{-1} \hat{L}^\top \hat{\Psi}^{-1} (x_j - \bar{x}), \quad j = 1, 2, \dots, n\end{aligned}\quad (30)$$

hoặc

$$\begin{aligned}\hat{f}_j &= (\hat{L}_z^\top \hat{\Psi}_z^{-1} \hat{L}_z)^{-1} \hat{L}_z^\top \hat{\Psi}_z^{-1} z_j \\ &= \hat{\Delta}_z^{-1} \hat{L}_z^\top \hat{\Psi}_z^{-1} z_j, \quad j = 1, 2, \dots, n\end{aligned}$$

Trong đó $z_j = D^{-1/2}$ và $\hat{\rho} = \hat{L}_z \hat{L}_z^\top + \hat{\Psi}_z$.

Điểm nhân tố được xác định trong công thức (30) có vector trung bình và ma trận hiệp phương sai bằng 0.

Phương pháp bình phương tối thiểu có trọng số

Nếu ma trận tải sau khi quay $\hat{L}^* = \hat{L}T$ được sử dụng để thay thế ma trận tải ban đầu thì các điểm nhân tố mới $\hat{f}_j^* = T^\top \hat{f}_j$, $j = 1, 2, \dots, n$.

Nếu ma trận tải được ước lượng bằng phương pháp thành phần chính, thông thường ta sẽ tạo điểm nhân tố bằng cách sử dụng bình phương tối thiểu không có trọng số. Một cách ngầm định, điều này có nghĩa rằng chúng giống nhau hoặc gần giống nhau. Khi đó các điểm nhân tố:

$$\hat{f}_j = (\tilde{L}^\top \tilde{L})^{-1} \tilde{L}^\top (x_j - \bar{x})$$

hoặc

$$\hat{f}_j = (\tilde{L}^\top \tilde{L})^{-1} \tilde{L}^\top z_j$$

với dữ liệu chuẩn hóa.

Phương pháp bình phương tối thiểu có trọng số

Vì $\tilde{L} = [\sqrt{\lambda_1}e_1 \mid \sqrt{\lambda_2}e_2 \mid \dots \mid \sqrt{\lambda_m}e_m]$ nên ta có:

$$\hat{f}_j = \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{e}_1^\top (x_j - \bar{x}) \\ \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{e}_2^\top (x_j - \bar{x}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}} \hat{e}_m^\top (x_j - \bar{x}) \end{bmatrix} \quad (31)$$

Phương pháp bình phương tối thiểu có trọng số

Với các điểm nhân tố này:

$$\frac{1}{n} \sum_{j=1}^n \hat{f}_j = 0 \quad (\text{trung bình mẫu})$$

$$\frac{1}{n-1} \sum_{j=1}^n \hat{f}_j \hat{f}_j^\top = I \quad (\text{phương sai mẫu hiệu chỉnh})$$

Phương pháp bình phương tối thiểu có trọng số

Ví dụ 9.5.1

Quay trở lại với ví dụ ?? về giá cổ phiếu, ta đã có ước lượng của ma trận tải sau khi quay và phương sai cụ thể bằng phương pháp ước lượng hợp lý cực đại:

$$\hat{L}_z^* = \begin{bmatrix} .763 & .024 \\ .821 & .227 \\ .669 & .104 \\ .118 & .993 \\ .113 & .675 \end{bmatrix} \quad \text{và} \quad \hat{\Psi}_z = \begin{bmatrix} .42 & 0 & 0 & 0 & 0 \\ 0 & .27 & 0 & 0 & 0 \\ 0 & 0 & .54 & 0 & 0 \\ 0 & 0 & 0 & .00 & 0 \\ 0 & 0 & 0 & 0 & .53 \end{bmatrix}$$

Phương pháp bình phương tối thiểu có trọng số

Vector của các quan sát đã chuẩn hóa:

$$z^{\top} = [.50 \quad -1.40 \quad -.20 \quad -.70 \quad 1.40]$$

Ta có điểm nhân tố cho các nhân tố 1 và 2:

Phương pháp bình phương tối thiểu có trọng số (30):

$$\hat{f} = (\hat{L}_z^{*\top} \hat{\Psi}_z^{-1} \hat{L}_z^*)^{-1} \hat{L}_z^{*\top} \hat{\Psi}_z^{-1} z = \begin{bmatrix} -.61 \\ -.61 \end{bmatrix}$$

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố**
 - Phương pháp bình phương tối thiểu có trọng số
 - Phương pháp hồi quy
- 6 Vận dụng

Phương pháp hồi quy

Ta xuất phát từ mô hình nhân tố ban đầu, coi như đã biết ma trận tải L và ma trận phương sai Ψ :

$$X - \mu = LF + \epsilon$$

Khi các nhân tố chung F và các nhân tố xác định ϵ có cùng phân phối chuẩn với trung bình và hiệp phương sai được cho bởi mô hình thì $X - \mu = LF + \epsilon$ có phân phối chuẩn $N_p(0, LL^T + \Psi)$.

Hơn nữa, phân phối chung của $(X - \mu)$ và F là $N_{m+p}(0, \Sigma^*)$, trong đó:

$$\Sigma^*_{(m+p) \times (m+p)} = \begin{bmatrix} \Sigma = \underset{(p \times p)}{LL^T} + \Psi & \underset{(p \times m)}{L} \\ \underset{(m \times p)}{L^T} & \underset{(m \times m)}{I} \end{bmatrix} \quad (32)$$

và 0 là vector không cỡ $(m + p) \times 1$.

Nhắc lại kết quả ở chương 4:

Cho $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ có phân phối chuẩn $N_p(\mu, \epsilon)$ với $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, và $|\Sigma_{22}| > 0$. Khi đó, phân phối có điều kiện của X_1 với điều kiện $X_2 = x_2$ là phân phối chuẩn và có:

$$\text{Trung bình} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\text{Hiệp phương sai} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Hơn nữa, hiệp phương sai này không phụ thuộc vào giá trị x_2 của biến điều kiện.

Phương pháp hồi quy

Sử dụng kết quả này, ta tìm được phân phối có điều kiện của $F|x$ là phân phối chuẩn đa biến với:

$$E(F|x) = L^T \Sigma^{-1}(x - \mu) = L^T (LL^T + \Psi)^{-1}(x - \mu) \quad (33)$$

$$\text{Cov}(F|x) = I - L^T \Sigma^{-1} L = I - L^T (LL^T + \Psi)^{-1} L \quad (34)$$

Các đại lượng $L^T (LL^T + \Psi)^{-1}$ là các hệ số hồi quy của các nhân tố trên các biến. Ước lượng của các hệ số này tạo ra các điểm nhân tố tương tự như ước lượng của các giá trị trung bình trong phân tích hồi quy đa biến.

Phương pháp hồi quy

Do đó, với bất kì vector quan sát được x_j và lấy các ước lượng hợp lí cực đại của \hat{L} và $\hat{\Psi}$ là các giá trị thực, ta thấy vector điểm nhân tố thứ j được xác định bởi:

$$\hat{f}_j = \hat{L}^\top \hat{\Sigma}^{-1}(x_j - \bar{x}) = \hat{L}^\top (\hat{L}\hat{L}^\top + \hat{\Psi})^{-1}(x_j - \bar{x}), \quad j = 1, 2, \dots, n \quad (35)$$

Kết quả của \hat{f}_j trong (35) có thể được rút gọn bằng cách sử dụng ma trận đơn vị:

$$\underset{(m \times p)}{\hat{L}^\top} \underset{(p \times p)}{(\hat{L}\hat{L}^\top + \hat{\Psi})^{-1}} = (I + \underset{(m \times m)}{\hat{L}^\top \hat{\Psi}^{-1} \hat{L}})^{-1} \underset{(m \times p)}{\hat{L}^\top} \underset{(p \times p)}{\hat{\Psi}^{-1}} \quad (36)$$

Phương pháp hồi quy

Ta có thể so sánh các điểm nhân tố trong công thức (35), được tạo bởi phương pháp hồi quy với các điểm nhân tố được tạo bởi phương pháp bình phương tối thiểu ở công thức (30). Ta kí hiệu điểm nhân tố được tạo bởi hai phương pháp lần lượt là \hat{f}_j^R và \hat{f}_j^{LS} . Từ công thức (36), ta có:

$$\hat{f}_j^{LS} = (\hat{L}^\top \hat{\Psi}^{-1} \hat{L})^{-1} (I + \hat{L}^\top \hat{\Psi}^{-1} \hat{L}) \hat{f}_j^R = (I + (\hat{L}^\top \hat{\Psi}^{-1} \hat{L})^{-1}) \hat{f}_j^R \quad (37)$$

Phương pháp hồi quy

Với các ước lượng hợp lí cực đại $(\hat{L}^\top \hat{\Psi}^{-1} \hat{L})^{-1} = \hat{\Delta}^{-1}$ và nếu các phần tử của ma trận đường chéo này gần bằng 0 thì hai phương pháp sẽ cho ra các điểm nhân tố gần bằng nhau.

Để giảm thiểu ảnh hưởng của một quyết định sai về số lượng các nhân tố, ta thường có xu hướng sử dụng ma trận S (ma trận hiệp phương sai mẫu ban đầu) thay cho $\hat{\Sigma} = \hat{L}\hat{L}^\top + \hat{\Psi}$. Do đó ta có kết quả sau:

Điểm nhân tố thu được bằng phương pháp hồi quy

$$\hat{f}_j = \hat{L}^\top S^{-1}(x_j - \bar{x}), \quad j = 1, 2, \dots, n \quad (38)$$

hoặc

$$\hat{f}_j = \hat{L}_z^\top R^{-1} z_j$$

Trong đó, $z_j = D^{-1/2}(x_j - \bar{x})$ và $\hat{\rho} = \hat{L}_z \hat{L}_z^\top + \hat{\Psi}_z$

Nếu ma trận tải sau khi quay $\hat{L}^* = \hat{L}T$ được sử dụng để thay thế ma trận tải ban đầu thì điểm nhân tố mới $\hat{f}_j = T^\top \hat{f}_j$, $j = 1, 2, \dots, n$.

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng**
 - Data 1
 - Data 2
 - Data 3

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng**
 - Data 1
 - Data 2
 - Data 3

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng**
 - Data 1
 - Data 2
 - Data 3

- 1 Đặt vấn đề
- 2 Mô hình phân tích nhân tố trực giao
- 3 Các phương pháp ước lượng
- 4 Xoay nhân tố
- 5 Điểm nhân tố
- 6 Vận dụng**
 - Data 1
 - Data 2
 - **Data 3**

Bảng đánh giá thành viên



ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

THANK YOU

NHÓM 2

