Factor Analysis

Peerapon S.

CPE 378 Machine Learning (2/64)

Data and codes : https://bit.ly/cpe378ML_HDS

คะแนน 10 วิชาของนักเรียน

Person	Math (X1)	French (X2)	...	Music (X10)
1	11.2	45		50.1
2	10.5	48		51.3
3	9.8	51		44.7
4	12.3	52		54.4
...				
999	10.3	46		60.2
1000	12.0	41		41.6

ปัจจัยที่วัดไม่ได้โดยตรง

ความฉลาด (Intelligence) ?

ทักษะทางกายภาพ (Physical Skills) ?

Person	Factor 1	Factor 2
1		
2		
3		
4		
...		
999		
1000		

Topics

- ☐ Factor model and fitting
- ☐ Number of factors and variance explained
- ☐ Factor rotation
- ☐ Factor scores

Performance in 10 events of the decathlon

Person	Event1 (X1)	Event2 (X2)	...	Event10 (X10)
1	11.2	45		50.1
2	10.5	48		51.3
3	9.8	51		44.7
4	12.3	52		54.4
...				
999	10.3	46		60.2
1000	12.0	41		41.6

ปัจจัยที่วัดไม่ได้โดยตรง

ความทนทาน ?

ความแข็งแรง ?

Person	Factor 1	Factor 2
1		
2		
3		
4		
...		
999		
1000		

What is Factor Analysis ?

- Express the original variables (measurable / observable) as linear combinations of a small number of (unobservable) hidden/latent variables, called ***factors***.
  - IQ test score; Intelligence and mental ability
  - A person does so well on standardized tests because of his/her high level of intelligence.
  - Athelete performance; Speed, Strength
- Unlike PCA, more understanding of the data could be achieved.
  - Latent variables typically used to give a formal representation of ideas or concepts that cannot be well-defined or measured directly.
  - General intelligence, verbal ability, mental ability, ambition, socioeconomic status, quality of life, happiness, etc.

Correlation of Two Standardized Variables

- สมมุติให้เรามีตัวแปรสุ่ม X_1, X_2 ที่มีค่าเฉลี่ยเป็นศูนย์และค่าเบี่ยงเบนมาตรฐานเท่ากับ 1
- ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X_1, X_2 มีค่าเป็น

$$\mathbb{C}\text{orr}\{X_1, X_2\} \triangleq \rho_{X_1,X_2} = \mathbb{E}\left\{\left(\frac{X_1-\mu_1}{\sigma_1}\right)\left(\frac{X_2-\mu_2}{\sigma_2}\right)\right\} = \mathbb{E}\{X_1X_2\}$$

Correlation Matrix

- สมมุติให้เรามีชุดตัวแปรสุ่ม p ตัว แทนด้วย $\mathbf{X} = [X_1, X_2, \ldots, X_p]$ โดยที่ X_i แต่ละตัวถูกทำให้เป็นมาตรฐานแล้ว (ลบด้วยค่าเฉลี่ย μ_i และหารด้วยค่าเบี่ยงเบนมาตรฐาน σ_i)
- กำหนดให้ ρ_{X_i,X_j} แทนด้วย ρ_{ij} จะได้ว่า เมทริกซ์สหสัมพันธ์ของตัวแปรมาตรฐาน (Standardized variables) p ตัว มีค่าเป็น

$$\mathbb{C}\mathrm{orr}\{\boldsymbol{X}\} \triangleq \boldsymbol{R} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{bmatrix} = \begin{bmatrix} \mathbb{E}\{X_1X_1\} & \mathbb{E}\{X_1X_2\} & \cdots & \mathbb{E}\{X_1X_p\} \\ \mathbb{E}\{X_2X_1\} & \mathbb{E}\{X_2X_2\} & \cdots & \mathbb{E}\{X_2X_p\} \\ \vdots & \vdots & & \vdots \\ \mathbb{E}\{X_pX_1\} & \mathbb{E}\{X_pX_2\} & \cdots & \mathbb{E}\{X_pX_p\} \end{bmatrix} = \mathbb{E}\{\boldsymbol{X}\boldsymbol{X}^T\}$$

ความแปรปรวนรวมในข้อมูลมาตรฐาน (standardized data) = p

Sample Correlation Matrix

- ถ้าเรามีกลุ่มตัวอย่างขนาด n ของชุดตัวแปรสุ่ม p ตัว แทนด้วยเวคเตอร์ $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$, $\mathbf{x}_i \in R^n$ โดยที่ $\mathbf{x}_i$ แต่ละตัวถูกทำให้เป็นมาตรฐานแล้ว
- เมทริกซ์สหสัมพันธ์ของตัวแปรมาตรฐาน p ตัว สามารถประมาณได้จากข้อมูลตัวอย่างในรูปแบบของเมทริกซ์สหสัมพันธ์ตัวอย่าง

$$\hat{\boldsymbol{R}} = \begin{bmatrix} \hat{\rho}_{11} & \hat{\rho}_{12} & \cdots & \hat{\rho}_{1p} \\ \hat{\rho}_{21} & \hat{\rho}_{22} & \cdots & \hat{\rho}_{2p} \\ \vdots & \vdots & & \vdots \\ \hat{\rho}_{p1} & \hat{\rho}_{p2} & \cdots & \hat{\rho}_{pp} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} \boldsymbol{x}_1 \cdot \boldsymbol{x}_1 & \boldsymbol{x}_1 \cdot \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1 \cdot \boldsymbol{x}_p \\ \boldsymbol{x}_2 \cdot \boldsymbol{x}_1 & \boldsymbol{x}_2 \cdot \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_2 \cdot \boldsymbol{x}_p \\ \vdots & \vdots & & \vdots \\ \boldsymbol{x}_p \cdot \boldsymbol{x}_1 & \boldsymbol{x}_p \cdot \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p \cdot \boldsymbol{x}_p \end{bmatrix}$$

$\mathbf{x}_1$ $\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_p$

	wheel-base	length	width	height	curb-weight	engine-size	compression-ratio	city-mpg	highway-mpg
0	88.6	168.8	64.1	48.8	2548	130	9.0	21	27
1	88.6	168.8	64.1	48.8	2548	130	9.0	21	27
2	94.5	171.2	65.5	52.4	2823	152	9.0	19	26
3	99.8	176.6	66.2	54.3	2337	109	10.0	24	30
4	99.4	176.6	66.4	54.3	2824	136	8.0	18	22
...	...	...	...	...	...	...	...	...	...
200	109.1	188.8	68.9	55.5	2952	141	9.5	23	28
201	109.1	188.8	68.8	55.5	3049	141	8.7	19	25
202	109.1	188.8	68.9	55.5	3012	173	8.8	18	23
203	109.1	188.8	68.9	55.5	3217	145	23.0	26	27
204	109.1	188.8	68.9	55.5	3062	141	9.5	19	25

$$\hat{\rho}_{12} = \frac{1}{n}\mathbf{x}_1 \cdot \mathbf{x}_2$$

$$= \frac{1}{205}(88.6 \cdot 168.8 + 88.6 \cdot 16 + 94.5 \cdot 171.2 + \cdots + 109.1 \cdot 188.8)$$

Constructing the Factor Model

- Observable random vector $\mathbf{X} = [X_1, X_2, \ \dots \ , X_p]$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$
- Express the standardized data in terms of the unobservable random variables $F_1, F_2, \ \dots \ , F_m$ (m common factors), $m \ll p$

$$
\begin{aligned}
\frac{X_1 - \mu_1}{\sigma_1} &= X_1^{(s)} = \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1 \\
\frac{X_2 - \mu_2}{\sigma_2} &= X_2^{(s)} = \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \varepsilon_2 \\
&\vdots \qquad\qquad \vdots \\
\frac{X_p - \mu_p}{\sigma_p} &= X_p^{(s)} = \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p
\end{aligned}
$$

$X_i^{(s)}$: Standardized X_i

ℓ_{ij}: Factor loading of ith variable on jth factor

F_i: Common factor having zero mean with unit variance

ε_i: Specific factor having zero mean with variance ψ_i

F_i, ε_i independent

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

- Deriving $\mathbf{L}, \mathbf{F}, \epsilon$ are not possible unless some assumptions were made.

Orthogonal Factor Model

Matrix of factor loadings

$$\underset{(p\times 1)}{\boldsymbol{X}} = \underset{(p\times m)}{\boldsymbol{L}}\underset{(m\times 1)}{\boldsymbol{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}, \quad m < p$$

$$\boldsymbol{X} = \begin{bmatrix} X_1^{(s)} \\ X_2^{(s)} \\ \vdots \\ X_p^{(s)} \end{bmatrix}, \quad \boldsymbol{L} = \begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{p1} & \ell_{p2} & \cdots & \ell_{pm} \end{bmatrix}, \quad \boldsymbol{F} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}, \quad \mathbb{V}\mathrm{ar}\{\varepsilon_i\} = \psi_i$$

$\boldsymbol{F}, \boldsymbol{\varepsilon}$ independent with

$$\mathbb{E}\{\boldsymbol{F}\} = \underset{(m\times 1)}{\boldsymbol{0}}, \quad \mathbb{C}\mathrm{ov}\{\boldsymbol{F}\} = \mathbb{E}\{\boldsymbol{F}\boldsymbol{F}^T\} = \underset{(m\times m)}{\boldsymbol{I}}$$

$$\mathbb{E}\{\boldsymbol{\varepsilon}\} = \underset{(p\times 1)}{\boldsymbol{0}}, \quad \mathbb{C}\mathrm{ov}\{\boldsymbol{\varepsilon}\} = \mathbb{E}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\} = \underset{(p\times p)}{\boldsymbol{\Psi}} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\mathbb{C}\text{orr}\{\boldsymbol{X}\} \triangleq \boldsymbol{R} = \mathbb{E}\{\boldsymbol{X}\boldsymbol{X}^T\} \quad \text{(Correlation Matrix)}$$

$$= \boldsymbol{L}\,\boldsymbol{L}^T + \boldsymbol{\Psi}$$

$$\boldsymbol{L} = \begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{p1} & \ell_{p2} & \cdots & \ell_{pm} \end{bmatrix}, \quad \boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

$$\mathbb{V}\text{ar}\{X_i^{(s)}\} = 1 = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2}_{\text{Communuality } h_i^2} + \underbrace{\psi_i}_{\text{Specific variance}}, \; |\ell_{ij}| \leq 1$$

$$= h_i^2 + \psi_i, \quad i = 1, 2, \ldots, p, \; \psi_i \geq 0$$

What is the total variance in data ?

Goal: หาค่า $\mathbf{L}$ กับ $\mathbf{\Psi}$ ที่ทำให้ $\mathbf{L}\,\mathbf{L}^T + \mathbf{\Psi}$ มีค่าใกล้เคียงกับ $\mathbf{R}$ มากที่สุด

Example: Ratings of Acquaintances

- เด็กนักเรียนได้เก็บคะแนนความพึงพอใจ 5 ด้าน "Kind" (ความใจดี), "Intelligent" (ความฉลาด), "Happy" (ความสนุกสนานร่าเริง), "Likeable" (ความน่าคบ), "Just" (ความมีเหตุมีผล) จากคนรู้จัก 7 คน ในระดับคะแนน 1 ถึง 9 ได้ผลลัพธ์ตามตารางด้านล่าง

People	Kind	Intelligent	Happy	Likable	Just
P1	1	5	5	1	1
P2	8	9	7	9	8
P3	9	8	9	9	8
P4	9	9	9	9	9
P5	1	9	1	1	9
P6	9	7	7	9	9
P7	9	7	9	9	7

```python
import pandas as pd

ratings_df = pd.DataFrame(data=[[1,5,5,1,1],
                                [8,9,7,9,8],
                                [9,8,9,9,8],
                                [9,9,9,9,9],
                                [1,9,1,1,9],
                                [9,7,7,9,9],
                                [9,7,9,9,7]],
                          columns = ['Kind','Intelligent','Happy','Likable','Just'],
                          index = ['P1','P2','P3','P4','P5','P6','P7'])
ratings_df.corr()
```

	Kind	Intelligent	Happy	Likable	Just
Kind	1.000	0.296	0.881	0.995	0.545
Intelligent	0.296	1.000	-0.022	0.326	0.837
Happy	0.881	-0.022	1.000	0.867	0.130
Likable	0.995	0.326	0.867	1.000	0.544
Just	0.545	0.837	0.130	0.544	1.000

ติดตั้งโดยใช้คำสั่ง pip install factor_analyzer

```
import numpy as np
from factor_analyzer import FactorAnalyzer

np.set_printoptions(precision=4)

fa = FactorAnalyzer(n_factors=2, method='ml', rotation=None, is_corr_matrix=False)


fa.fit(ratings_df)
print(fa.loadings_)
```

```
[[ 0.9687 -0.2308]
 [ 0.519   0.8056]
 [ 0.7839 -0.5863]
 [ 0.9701 -0.2096]
 [ 0.7034  0.6658]]
```

$$X_1 = .9687\,F_1 - .2308\,F_2$$
$$X_2 = .519\,F_1 + .8056\,F_2$$
$$X_3 = .7839\,F_1 - .5863\,F_2$$
$$X_4 = .9701\,F_1 - .2096\,F_2$$
$$X_5 = .7034\,F_1 + .6658\,F_2$$

Factor loadings

$$
\begin{aligned}
\boldsymbol{L}\,\boldsymbol{L}^T + \boldsymbol{\Psi} &= \begin{pmatrix} .969 & -.231 \\ .519 & .807 \\ .785 & -.587 \\ .971 & -.210 \\ .704 & .667 \end{pmatrix} \begin{pmatrix} .969 & .519 & .785 & .971 & .704 \\ -.231 & .807 & -.587 & -.210 & .667 \end{pmatrix} \\
&\quad + \begin{pmatrix} .007 & 0 & 0 & 0 & 0 \\ 0 & .079 & 0 & 0 & 0 \\ 0 & 0 & .040 & 0 & 0 \\ 0 & 0 & 0 & .013 & 0 \\ 0 & 0 & 0 & 0 & .060 \end{pmatrix} \qquad \text{Specfic variances} \\
&= \begin{pmatrix} 1.000 & .317 & .896 & .990 & .528 \\ .317 & 1.000 & -.066 & .335 & .904 \\ .896 & -.066 & 1.000 & .885 & .161 \\ .990 & .335 & .885 & 1.000 & .543 \\ .528 & .904 & .161 & .543 & 1.000 \end{pmatrix},
\end{aligned}
$$

- ค่า Loadings ℓ_{ij} ที่ได้ขึ้นอยู่กับวิธีที่ใช้ในการหาคำตอบ เช่น วิธี Maximum likelihood หรือ วิธี Principal factor เป็นต้น
- ค่า Loadings ℓ_{ij} แสดงระดับความสัมพันธ์ระหว่างตัวแปร X_i กับปัจจัย F_j

Observed variables	Loadings F1	Loadings F2
Kind	0.9687	-0.2308
Intelligent	0.5190	0.8056
Happy	0.7839	-0.5863
Likable	0.9701	-0.2096
Just	0.7034	0.6658

$$X_1 = .9687\,F_1 - .2308\,F_2$$
$$X_2 = .519\,F_1 + .8056\,F_2$$
$$X_3 = .7839\,F_1 - .5863\,F_2$$
$$X_4 = .9701\,F_1 - .2096\,F_2$$
$$X_5 = .7034\,F_1 + .6658\,F_2$$

Self Test: Communualities and Explained Variances

Proportion of Variance explained by Factors

- สัดส่วนของความแปรปรวนของข้อมูลที่อธิบายโดยปัจจัย m ตัว สามารถประมาณได้จากผลบวกของค่าลักษณะเฉพาะ (Eigenvalue) m ตัวแรกของเมทริกซ์สหสัมพันธ์ $\mathbf{R}$
- ผลรวมของค่าลักษณะเฉพาะทั้งหมดมีค่าเท่ากับ p (จำนวน Observed Variables)

```
import numpy as np

eigval,eigvec=np.linalg.eig(ratings_df.corr())

# Sort descendingly
print(-np.sort(-eigval))
```

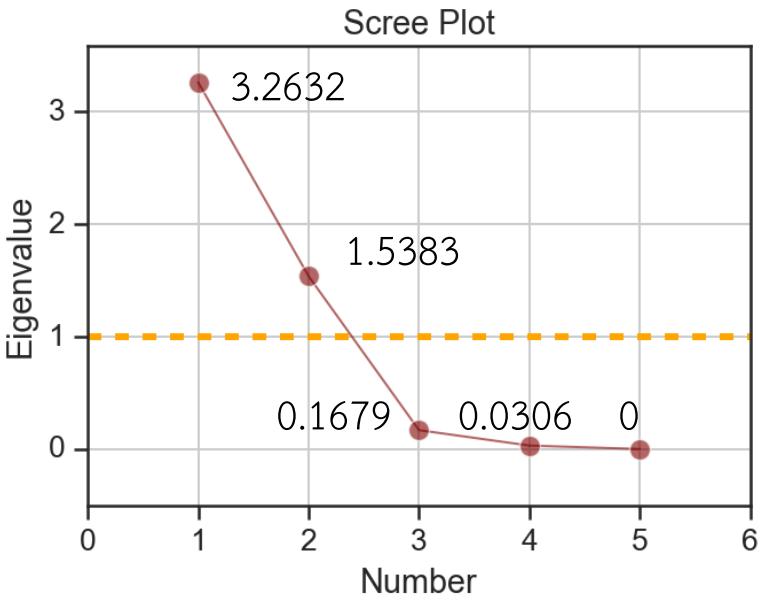
```
[ 3.2632  1.5383  0.1679  0.0306 -0.]
```

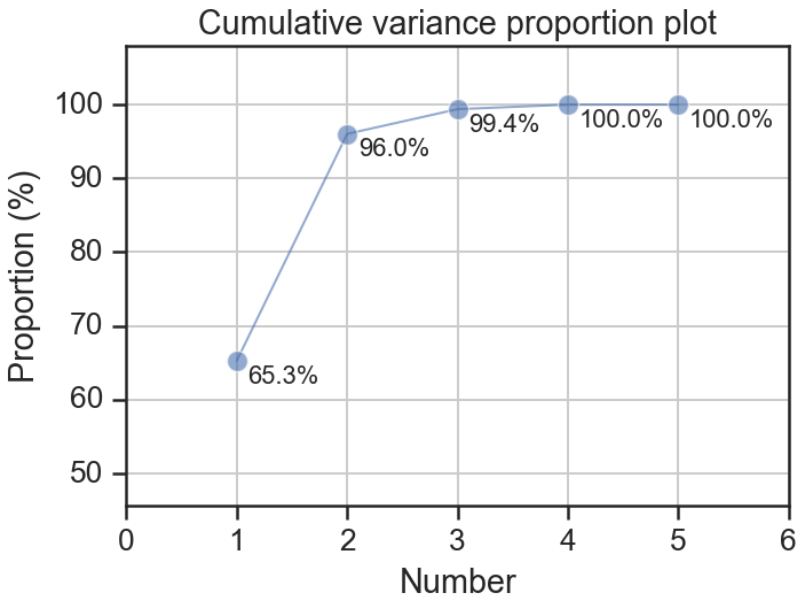
Observed variables	Loadings F1	Loadings F2
Kind	0.9687	-0.2308
Intelligent	0.5190	0.8056
Happy	0.7839	-0.5863
Likable	0.9701	-0.2096
Just	0.7034	0.6658
Total contributed variance	3.2582	1.5333

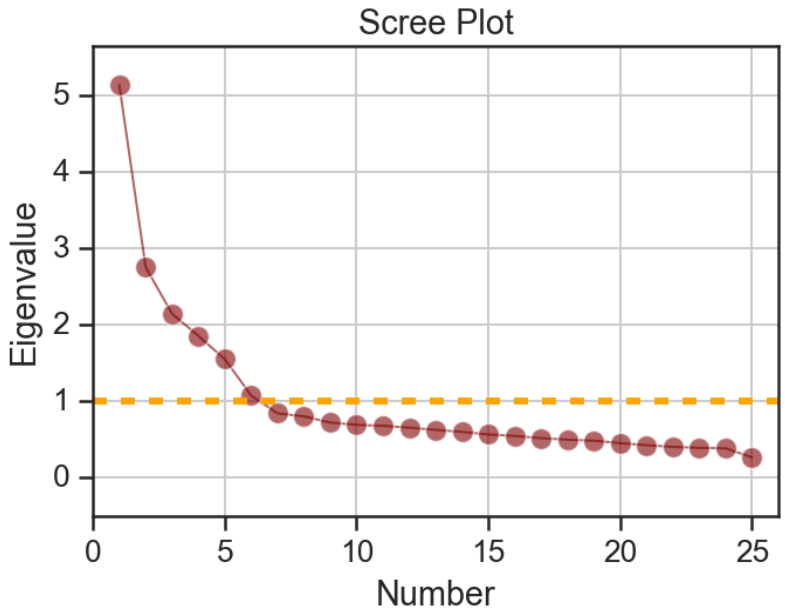
สัดส่วนของความแปรปรวนที่อธิบายโดยปัจจัยที่ 1 และ 2 รวมกัน $= \frac{3.2632 + 1.5383}{5} = 96\,\%$

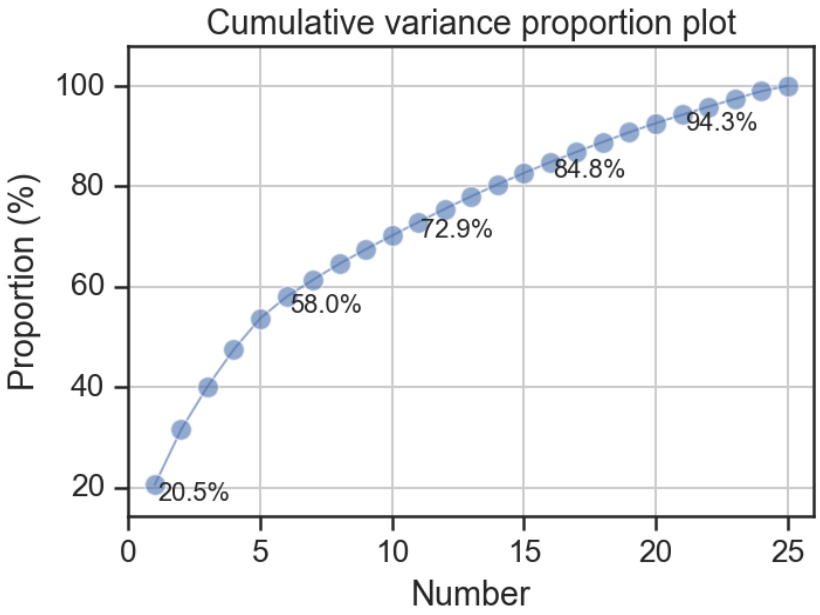
Choosing the number of factors

- ☐ แบบที่ 1: เลือกจำนวนปัจจัยเท่ากับจำนวนของค่าลักษณะเฉพาะที่มีค่ามากกว่า 1
- ☐ แบบที่ 2: ดูจาก Scree Plot
- ☐ แบบที่ 3: ดูจากสัดส่วนรวมของค่าลักษณะเฉพาะ (Cumulative variance proportion)
- ☐ อธิบายความแปรปรวนในข้อมูลให้ได้มากที่สุด (อย่างน้อย 50%)
  โดยใช้จำนวนปัจจัยไม่มากเกินไปจนไม่สามารถตีความได้

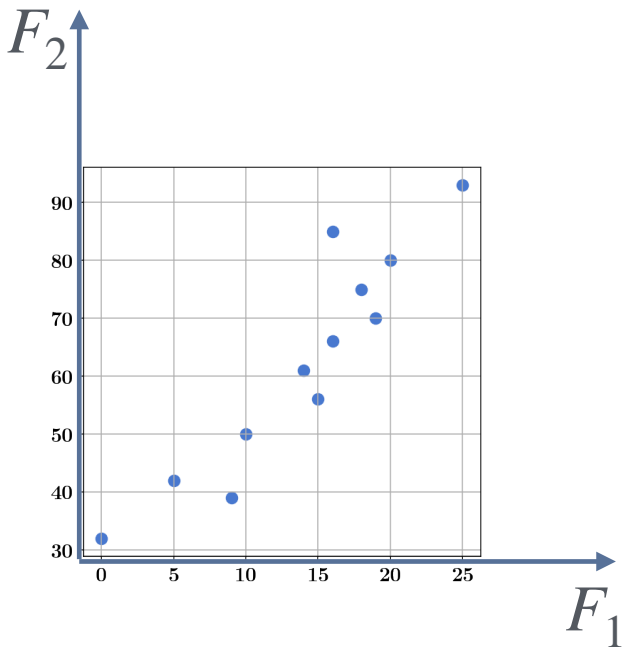

Scree Plot
3.2632
1.5383
0.1679
0.0306
0
Eigenvalue
Number
0
1
2
3
4
5
6

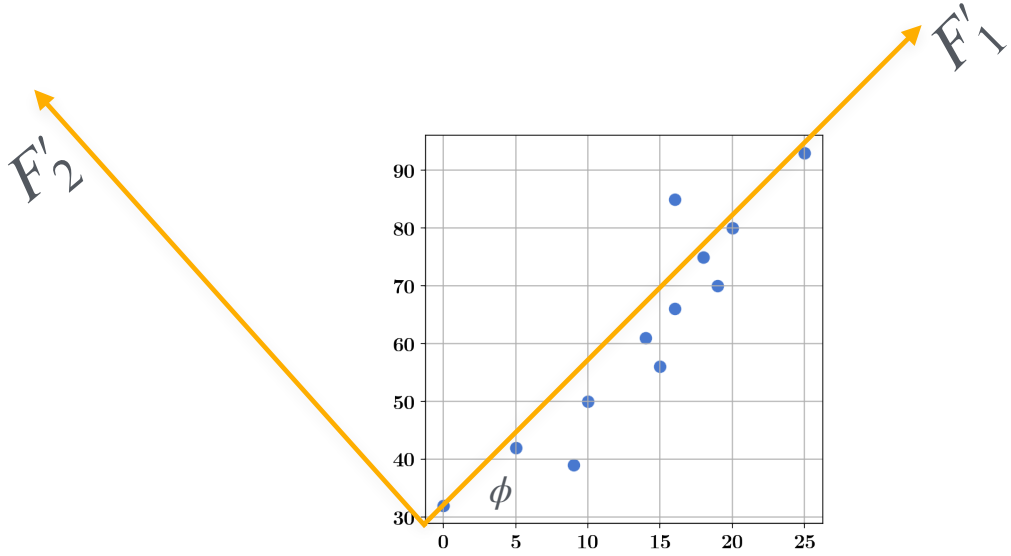

Cumulative variance proportion plot
65.3%
96.0%
99.4%
100.0%
100.0%
Proportion (%)
Number
50
60
70
80
90
100
0
1
2
3
4
5
6


Scree Plot
Eigenvalue
Number
0
1
2
3
4
5
0
5
10
15
20
25


Cumulative variance proportion plot
Proportion (%)
Number
20
40
60
80
100
0
5
10
15
20
25
20.5%
58.0%
72.9%
84.8%
94.3%

Factor Rotation





$$[F_1' \; F_2'] = [F_1 \; F_2] \underbrace{\begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}}_{\boldsymbol{T}}$$

Observed variables	Loadings F1	Loadings F2	Communualities	Specific var.
Kind	0.9687	-0.2308	0.9917	0.0083
Intelligent	0.5190	0.8056	0.9184	0.0816
Happy	0.7839	-0.5863	0.9583	0.0417
Likable	0.9701	-0.2096	0.9851	0.0149
Just	0.7034	0.6658	0.9382	0.0618

ค่า Loading Matrix ดั้งเดิม

```
fa = FactorAnalyzer(n_factors=2,
                    method='ml',
                    rotation=None)
```

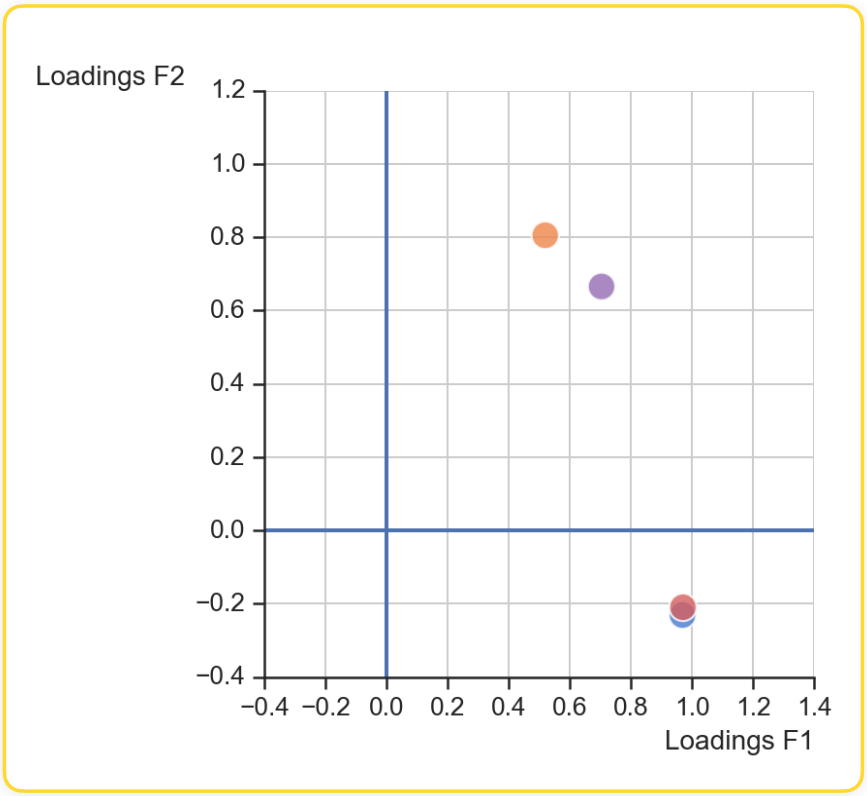
Observed variables	Loadings F1	Loadings F2	Communualities	Specific var.
Kind	0.9497	0.2995	0.9917	0.0083
Intelligent	0.0316	0.9578	0.9184	0.0816
Happy	0.9737	-0.1005	0.9583	0.0417
Likable	0.9401	0.3184	0.9851	0.0149
Just	0.2616	0.9326	0.9382	0.0618

ค่า Loading Matrix ที่ถูกหมุนแล้ว

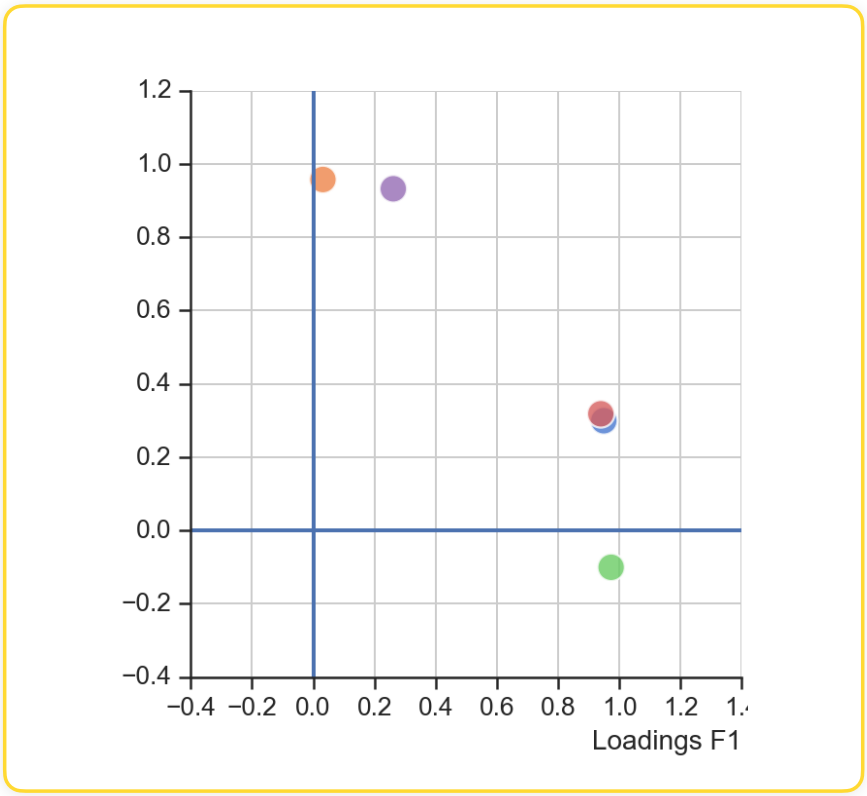
```
fa = FactorAnalyzer(n_factors=2,
                    method='ml',
                    rotation='varimax')
```

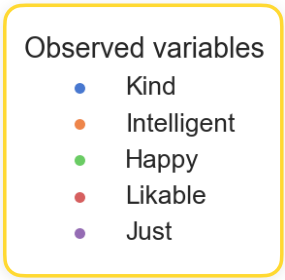
ให้ค่า Variance ของ Loadings ในแต่ละปัจจัยที่ค่าสูงสุด

No Rotation



'Varimax' Rotation



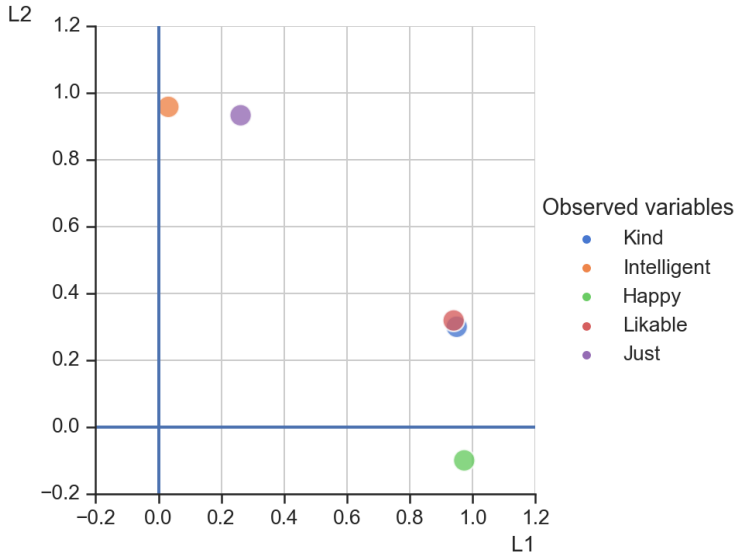


Self Test: Factor Rotation

Factor Analysis Result Interpretation

- Group together variables whose loadings of the same factors are high
  - Group 1: Kind, Happy, Likeable
  - Group 2: Intelligent, Just

Observed variables	Loadings F1	Loadings F2	Communualities	Specific var.
Kind	0.9497	0.2995	0.9917	0.0083
Intelligent	0.0316	0.9578	0.9184	0.0816
Happy	0.9737	-0.1005	0.9583	0.0417
Likable	0.9401	0.3184	0.9851	0.0149
Just	0.2616	0.9326	0.9382	0.0618



Factor Scores

- จากค่าตามชุดค่าตัวแปรที่สังเกตได้ $\mathbf{x} = [x_1, x_2, \ \dots \ , x_p]$ และค่าของ Loadings ℓ_{ij} เราสามารถหาค่าของปัจจัย m ตัว $\mathbf{f} = [f_1, f_2, \ \dots \ , f_m]$ สำหรับข้อมูลตัวอย่างนี้ได้
  - ใช้บอกพฤติกรรมของข้อมูลตัวอย่างนี้จากมุมมองของค่าปัจจัย
  - ใช้ลดมิติข้อมูลจาก p คอลัมน์ เหลือ m คอลัมน์

$$X_1^{(s)} = \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1$$

$$X_2^{(s)} = \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \varepsilon_2$$

$$\vdots \qquad\qquad \vdots$$

$$X_p^{(s)} = \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p$$

	x_1	x_2	...	x_p
1	11.2	45		50.1
2	10.5	48		51.3
3	9.8	51		44.7
...				
n	10.3	46		60.2

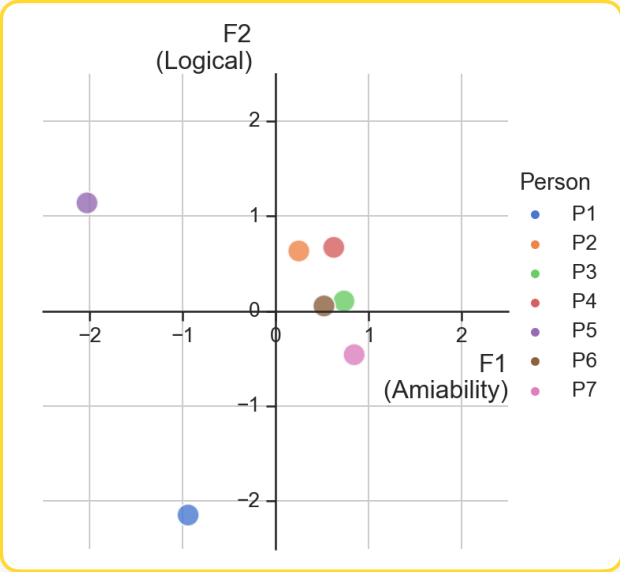
	f_1	f_2
1		
2		
3		
...		
n		

- Compute factor scores (only when the raw data is available)

```
fa.transform(ratings_df)
```

```
array([[-0.9423, -2.1462],
       [ 0.2488,  0.6333],
       [ 0.7347,  0.1063],
       [ 0.625 ,  0.6712],
       [-2.0304,  1.1394],
       [ 0.5198,  0.0546],
       [ 0.8444, -0.4585]])
```

People	Kind	Intelligent	Happy	Likable	Just
P1	1	5	5	1	1
P2	8	9	7	9	8
P3	9	8	9	9	8
P4	9	9	9	9	9
P5	1	9	1	1	9
P6	9	7	7	9	9
P7	9	7	9	9	7



Niether amiable nor logical

Logical but less amiabilty

Amiability but less logical

Good at both

Not amiability but logical

Neutral

Amiability but illogical

Person	F1	F2
P1	-0.9423	-2.1462
P2	0.2488	0.6333
P3	0.7347	0.1063
P4	0.6250	0.6712
P5	-2.0304	1.1394
P6	0.5198	0.0546
P7	0.8444	-0.4585

Assignment (Group of five or less): Customer Preference Data

- ☐ (25 minutes) Submit to LEB2