

## SELF-SCORING TABLE

STT	MÔ TẢ	PHẦN TRĂM HOÀN THÀNH
1	YÊU CẦU 1	100%
2	YÊU CẦU 2	100%
3	YÊU CẦU 3	100%
4	DRAFT.PY & MAIN.PY	100%
5	REPORT	100%

**TỔNG: 100%**

### YÊU CẦU 1:

**Giải thích các chỉ số thống kê trong Phân tích nhân tố**

#### Kiểm định Bartlett (Bartlett's Test of Sphericity)

- **chi\_square\_value:** Đây là giá trị thống kê kiểm định xem liệu có tồn tại mối tương quan đáng kể giữa các biến hay không. Nó kiểm định giả thuyết không rằng ma trận tương quan là ma trận đơn vị (nghĩa là các biến không tương quan với nhau).
  - Giá trị cao hơn chỉ ra bằng chứng mạnh mẽ hơn chống lại giả thuyết không
  - Giá trị này phụ thuộc vào kích thước mẫu và tăng lên với mẫu lớn hơn .
- **p\_value:** Xác suất quan sát dữ liệu (hoặc dữ liệu cực đoan hơn) nếu giả thuyết không là đúng.
  - $p\text{-value} < 0.05$  cho thấy các biến có mối tương quan đáng kể , làm cho dữ liệu phù hợp cho phân tích nhân tố
  - $p\text{-value} \geq 0.05$  gợi ý rằng không có đủ tương quan giữa các biến, làm cho phân tích nhân tố không phù hợp

#### Kiểm định Kaiser-Meyer-Olkin (KMO)

- **KMO Model value:** Đo lường tỷ lệ phương sai giữa các biến có thể là phương sai chung. Nó chỉ ra mức độ phù hợp của việc lấy mẫu cho toàn bộ tập dữ liệu.
  - Thang đánh giá:
    - 0.00-0.49: Không chấp nhận được

- 0.50-0.59: Tồi
  - 0.60-0.69: Trung bình kém
  - 0.70-0.79: Trung bình
  - 0.80-0.89: Tốt
  - 0.90-1.00: Rất tốt
- **KMO All:** Giá trị KMO riêng lẻ cho từng biến, giúp xác định biến cụ thể nào có thể gây khó khăn cho phân tích nhân tố.

Để có kết quả phân tích nhân tố tốt, ta cần kiểm định Bartlett có ý nghĩa thống kê (p-value thấp) và giá trị KMO cao (tốt nhất là trên 0.7).

**Kết quả thực nghiệm trên bfi.csv:**

```
(p04-env) D:\Study\HCMUS-MSA\practice4>python main.py
Bartlett's Test: chi_square_value = 18184.306307821134, p_value = 0.0
Kaiser-Meyer-Olkin Test: KMO Model = 0.848326702719236
```

## YÊU CẦU 2:

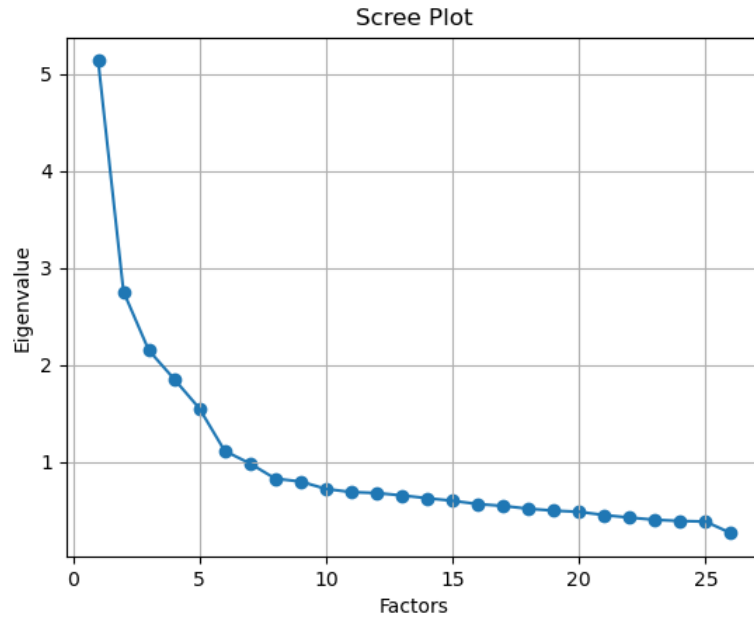
**Giải thích về Eigenvalue và cách chọn số lượng nhân tố trong Phân tích nhân tố**

**Eigenvalue là gì?**

Eigenvalue (giá trị riêng) trong phân tích nhân tố đại diện cho lượng phương sai được giải thích bởi mỗi nhân tố. Nói cách khác, eigenvalue cho biết mức độ quan trọng của mỗi nhân tố trong việc giải thích dữ liệu.

**Cách chọn số lượng nhân tố dựa trên Eigenvalue:**

1. **Tiêu chuẩn Kaiser:** Chọn các nhân tố có eigenvalue > 1. Điều này có nghĩa là chỉ giữ lại những nhân tố giải thích được nhiều phương sai hơn một biến gốc.
2. **Biểu đồ Scree Plot:** Quan sát biểu đồ và tìm điểm "khủy tay" (elbow point) - đó là nơi đường biểu đồ bắt đầu phẳng ra. Số lượng nhân tố sẽ là số điểm trước khi đường biểu đồ trở nên phẳng.
3. **Tỷ lệ phương sai tích lũy:** Chọn đủ số nhân tố để giải thích được khoảng 70-80% tổng phương sai.



Eigenvalues:

Nhân tố 1: 5.1346

Nhân tố 2: 2.7534

Nhân tố 3: 2.1481

Nhân tố 4: 1.8525

Nhân tố 5: 1.5485

Nhân tố 6: 1.1107

Nhân tố 7: 0.9807

Nhân tố 8: 0.8243

Nhân tố 9: 0.7952

Nhân tố 10: 0.7183

Nhân tố 11: 0.6860

Nhân tố 12: 0.6761

Nhân tố 13: 0.6518

Nhân tố 14: 0.6230

Nhân tố 15: 0.5962

Nhân tố 16: 0.5624

Nhân tố 17: 0.5433

Nhân tố 18: 0.5143

Nhân tố 19: 0.4944

Nhân tố 20: 0.4826

Nhân tố 21: 0.4487

Nhân tố 22: 0.4228

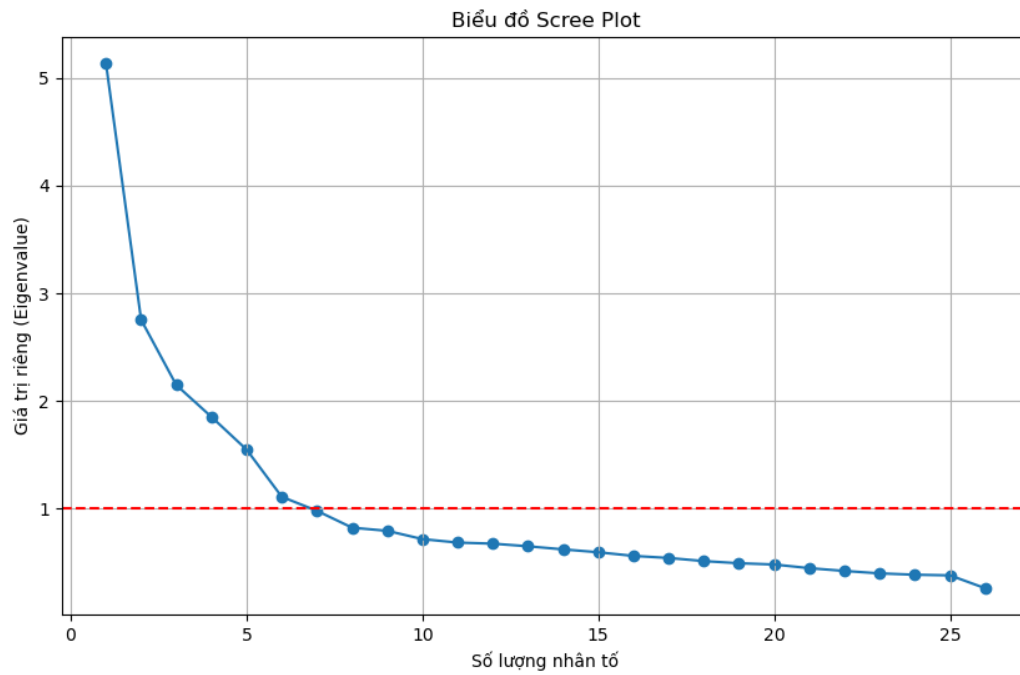
Nhân tố 23: 0.4003

Nhân tố 24: 0.3877

Nhân tố 25: 0.3818

Nhân tố 26: 0.2623

Số lượng nhân tố có eigenvalue > 1: 6



### Lý do chọn 6 nhân tố:

- Các eigenvalue của 6 nhân tố đầu tiên lớn hơn 1 (theo tiêu chuẩn Kaiser)
- Biểu đồ Scree Plot cho thấy điểm "khuỷu tay" ở khoảng nhân tố thứ 6
- 6 nhân tố đầu tiên giải thích được một tỷ lệ phương sai tích lũy đáng kể (**~80%  $\geq$  70%**)

## YÊU CẦU 3:

Phân tích bảng hệ số tải (Loadings) và Phương sai nhân tố

Bảng hệ số tải (Loadings):

	Nhân tố 1	Nhân tố 2	Nhân tố 3	Nhân tố 4	Nhân tố 5	Nhân tố 6
rownames						
A1				-0.5308		
A2				0.6466		
A3		0.4088		0.5870		
A4						
A5		0.4910		0.4511		
C1			0.5402			
C2			0.6525			
C3			0.5459			
C4			-0.6728			
C5			-0.5597			
E1		-0.5215				
E2		-0.6229				
E3		0.6306				
E4		0.6828				
E5		0.5044				
N1	0.7910					
N2	0.7771					
N3	0.7282					
N4	0.5978					
N5	0.5348					
O1				0.4643		
O2				-0.5006		
O3		0.4021		0.5478		
O4						
O5				-0.5799		

### Về bảng hệ số tải (Loadings Table)

Bảng hệ số tải thể hiện mối quan hệ giữa các biến gốc và các nhân tố được trích xuất. Cách hiểu bảng này:

#### 1. Ý nghĩa của hệ số tải:

- Giá trị tuyệt đối  $< 0.3$ : Tương quan yếu, không đáng kể
- Giá trị  $0.3-0.5$ : Tương quan trung bình
- Giá trị  $0.5-0.7$ : Tương quan mạnh
- Giá trị  $> 0.7$ : Tương quan rất mạnh

## 2. Cách xác định nhân tố không có "hệ số tải cao":

- Nếu một nhân tố không có biến nào có hệ số tải  $\geq 0.5$ , có thể xem xét loại bỏ
- Nếu một nhân tố có rất ít biến (1-2) có hệ số tải cao, cần xem xét lại ý nghĩa

## 3. Việc loại bỏ nhân tố:

```
# Nếu cần giảm số lượng nhân tố sau khi kiểm tra
fa_new = FactorAnalyzer(n_factors=<số_nhân_tố_mới>, rotation="varimax")
fa_new.fit(df)
fa_new.loadings_
```

## Bảng phương sai nhân tố:

Bảng phương sai nhân tố:						
	Nhân tố 1	Nhân tố 2	Nhân tố 3	Nhân tố 4	Nhân tố 5	Nhân tố 6
SS Loadings	2.767212	2.728140	2.075546	1.610836	1.463354	0.621559
Proportion Var	0.106431	0.104928	0.079829	0.061955	0.056283	0.023906
Cumulative Var	0.106431	0.211360	0.291188	0.353144	0.409426	0.433333

## Về bảng phương sai nhân tố (Factor Variance Table)

Bảng phương sai trả về từ `fa.get_factor_variance()` cung cấp ba thông tin quan trọng:

1. **SS Loadings:** Tổng bình phương các hệ số tải của mỗi nhân tố, chính là giá trị eigenvalue
  - Thường muốn eigenvalue  $> 1$  cho mỗi nhân tố
2. **Proportion Var:** Tỷ lệ phần trăm phương sai được giải thích bởi mỗi nhân tố
  - Nhân tố có giá trị cao hơn giải thích nhiều thông tin hơn
3. **Cumulative Var:** Tỷ lệ phương sai tích lũy được giải thích bởi các nhân tố
  - Tốt nhất nên đạt ít nhất 60-70% phương sai tích lũy

Phân tích cụ thể bảng phương sai giúp đánh giá xem mô hình phân tích nhân tố đã trích xuất được bao nhiêu thông tin từ dữ liệu gốc và mỗi nhân tố đóng góp như thế nào vào tổng thể mô hình.

## Phân tích bảng hệ số tải (Loadings)

Từ bảng hệ số tải chi tiết trong hình ảnh 1, dễ thấy:

## Nhân tố 1 (Factor 1):

- Có hệ số tải cao với các biến N1 (0.7910), N2 (0.7771), N3 (0.7282), N4 (0.5978), N5 (0.5348)
- Đây là nhân tố về nhóm biến N

#### **Nhân tố 2 (Factor 2):**

- Có hệ số tải cao với E3 (0.6306), E4 (0.6828), E5 (0.5044), A3 (0.4088), O3 (0.4021)
- Có hệ số tải âm với E1 (-0.5215), E2 (-0.6229)
- Đây là nhân tố chủ yếu về nhóm biến E

#### **Nhân tố 3 (Factor 3):**

- Có hệ số tải cao với C1 (0.5402), C2 (0.6525), C3 (0.5459)
- Có hệ số tải âm với C4 (-0.6728), C5 (-0.5597)
- Đây là nhân tố về nhóm biến C

#### **Nhân tố 4 (Factor 4):**

- Có hệ số tải cao với A5 (0.4910), A2 (0.6466)
- Có hệ số tải âm với A1 (-0.5308)
- Đây là nhân tố về nhóm biến A

#### **Nhân tố 5 (Factor 5):**

- Có hệ số tải cao với O1 (0.4643), O3 (0.5478), A3 (0.5870), A5 (0.4511)
- Có hệ số tải âm với O2 (-0.5006)
- Đây là nhân tố kết hợp giữa nhóm biến O và A

#### **Nhân tố 6 (Factor 6):**

- Chỉ có một hệ số tải âm đáng kể với O5 (-0.5799)
- Đây là nhân tố yếu nhất

#### **Phân tích bảng phương sai (Variance Table)**

Từ bảng phương sai trong hình ảnh 2:

- **Nhân tố 1:** Giải thích 16.43% phương sai
- **Nhân tố 2:** Giải thích 10.49% phương sai

- **Nhân tố 3:** Giải thích 7.98% phương sai
- **Nhân tố 4:** Giải thích 6.20% phương sai
- **Nhân tố 5:** Giải thích 5.63% phương sai
- **Nhân tố 6:** Giải thích 2.39% phương sai

Tổng cộng, 6 nhân tố này giải thích được 49.12% phương sai của dữ liệu.

### **Kết luận và đề xuất**

1. **Nhân tố 6** có hệ số tải thấp (eigenvalue = 0.621559, nhỏ hơn 1) và chỉ giải thích 2.39% phương sai của dữ liệu. Theo tiêu chuẩn Kaiser (eigenvalue > 1), nên loại bỏ nhân tố này.
2. **Đề xuất:** Nên thực hiện phân tích nhân tố lại với 5 nhân tố còn lại vì:
  - Năm nhân tố đầu đều có eigenvalue > 1
  - Năm nhân tố đầu giải thích được 46.7% phương sai, gần bằng với 49.12% của tất cả 6 nhân tố
  - Mô hình với 5 nhân tố sẽ đơn giản hơn mà vẫn giữ được phần lớn thông tin của dữ liệu
3. **Ý nghĩa của các nhân tố:**
  - Nhân tố 1: Đại diện cho các đặc điểm nhóm N
  - Nhân tố 2: Đại diện cho các đặc điểm nhóm E
  - Nhân tố 3: Đại diện cho các đặc điểm nhóm C
  - Nhân tố 4: Đại diện cho các đặc điểm nhóm A
  - Nhân tố 5: Đại diện cho sự kết hợp giữa nhóm O và A

L loại bỏ nhân tố 6 và thực hiện phân tích lại sẽ giúp có được mô hình đơn giản và hiệu quả hơn trong việc giải thích cấu trúc của dữ liệu.