Exploratory Factor Analysis (EFA)

Konstantinos I. Bougioukas, MSc, PhD

School of Psychology
Faculty of Philosophy
Aristotle University of Thessaloniki

Exploratory Factor Analysis

Exploratory Factor Analysis (EFA) focuses on identifying unobserved **latent variables** (also known as common factors) that explain the correlations among observed variables, with the aim of uncovering the data's underlying structure.

- **Observed variables** (also called manifest variables) are the variables that we can directly measure or observe.
- **Latent variables** are hidden or abstract concepts inferred from manifest variables.

Exploratory Factor Analysis (EFA) can be used to explore the dimensionality of a measurement instrument by identifying the smallest number of interpretable factors that explain the **common variance** among variables.

Exploratory Factor Analysis model

The factor model for the i-th observed variable (X_i) for a single individual can be expressed as a **regression on the common factors**:

Observed variable

$$X_i = l_{i1}F_1 + l_{i2}F_2 + \ldots + l_{im}F_k + u_i$$

Factor loading

Common factors

the unique factor associated with the observed variable

Steps in the process of EFA

1. Prepare the Data (standardization, outliers)
2. Evaluate Assumptions (Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity, multicollinearity or singularity)
3. Choose the Extraction Method (e.g., Principal Axis Factoring, Maximum Likelihood, or Maximum Residuals).
4. Determine the Number of Factors (eigenvalues greater than one, scree plot inspection, or parallel analysis).
5. Factor Rotation (Orthogonal Vs Oblique)
6. Interpret the Factors
7. Assess Factor Reliability (Cronbach's Alpha >0.7)
8. Refine the Model
9. Finalize the Factor Solution

Example of GRMS Stress Appraisal

25 Items

- Unattractive because of size of butt (Obj1)
- Negative comments about size of facial features (Obj2)
- Imitated the way they think Black women speak (Obj3)
- Someone made me feel unattractive (Obj4)
- Negative comment about skin tone (Obj5)
- Someone assumed I speak a certain way (Obj6)
- Objectified me based on physical features(Obj7)
- Someone assumed I have a certain body type (Obj8)
- Made a sexually inappropriate comment (Obj9)
- Negative comments about my hair when natural (Obj10)
- I have felt unheard (Marg1)
- My comments have been ignored (Marg2)
- . . .

EFA Assumptions

Identity matrix

$$I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- **Bartlett's test of sphericity**

A significant result ($p < 0.05$) indicates that the correlation matrix significantly differs from an identity matrix. This suggests that the variables share enough correlation to justify the use of principal component analysis (PCA).

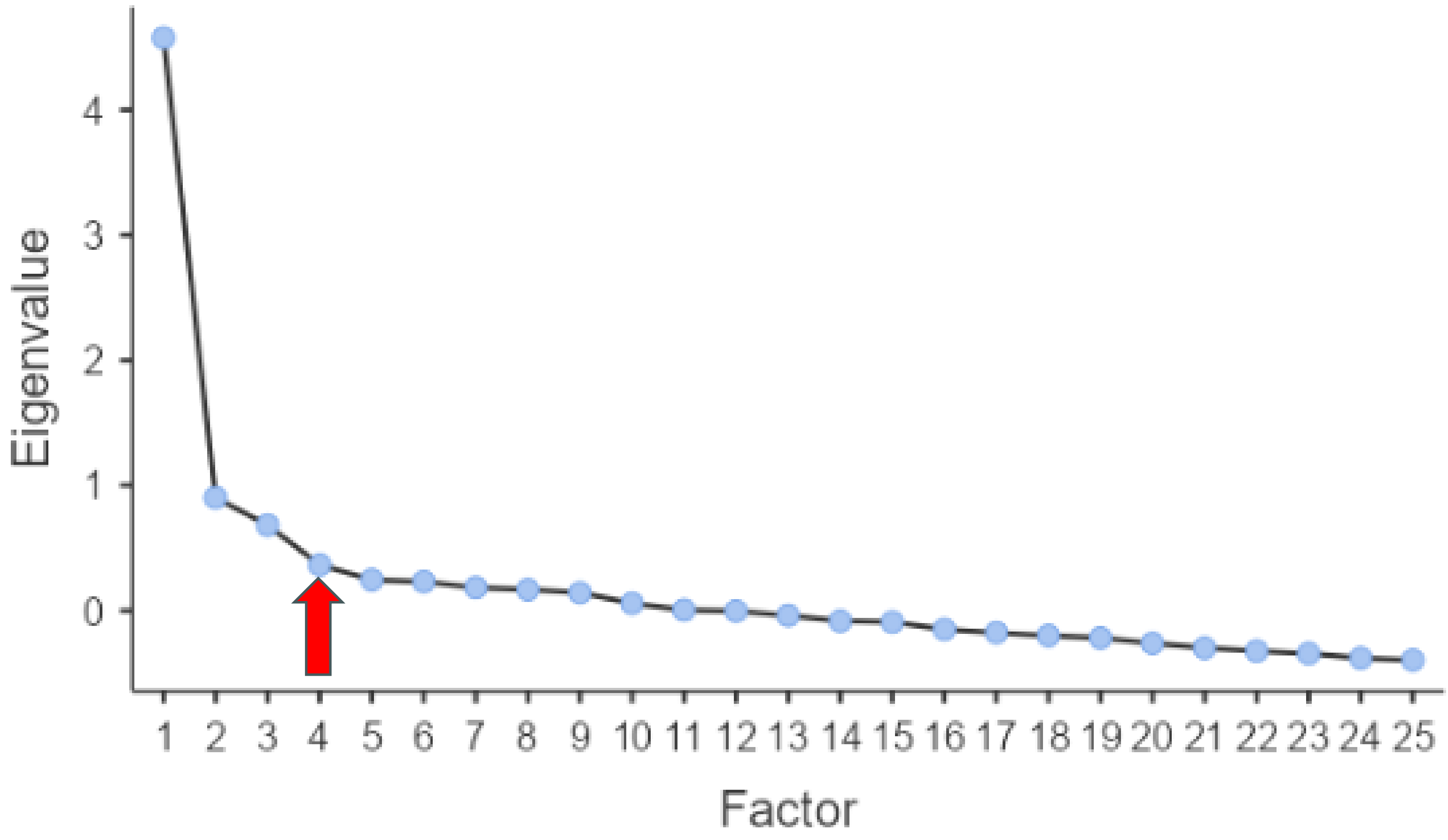
- **No multicollinearity and singularity**

- **Kaiser-Meyer-Olkin (MKO) index of Sampling Adequacy**

(bare minimum of 0.5, values between 0.5 and 0.7 as mediocre, values between 0.7 and 0.8 as good, values between 0.8 and 0.9 as great, values above 0.9 are superb)

- **Address outliers.**

Scree Plot-Number of factors



Pattern matrix

Factor loadings

Observed variables (features)

Loadings

Factor Loadings

	Factor				
	1	2	3	4	Uniqueness
Obj1	0.51				0.71
Obj2	0.53				0.68
Obj3	0.46				0.75
Obj4	0.56				0.67
Obj5	0.47				0.72
Obj6				0.25	0.78
Obj7	0.38				0.73
Obj8	0.53				0.65
Obj9	0.39				0.81
Obj10	0.48				0.78
Marg1		0.81			0.38
Marg2		0.49			0.59
Marg3		0.43			0.72
Marg4		0.40			0.72
Marg5		0.45			0.67
Marg6	0.31	0.25			0.73
Marg7		0.36			0.82
Str1			0.56		0.64
Str2			0.27		0.82
Str3			0.53		0.72
Str4			0.39		0.83
Str5			0.35		0.82
Ang1			0.33	0.24	0.78
Ang2				0.64	0.57
Ang3				0.30	0.81

Note. 'Principal axis factoring' extraction method was used in combination with a 'oblimin' rotation

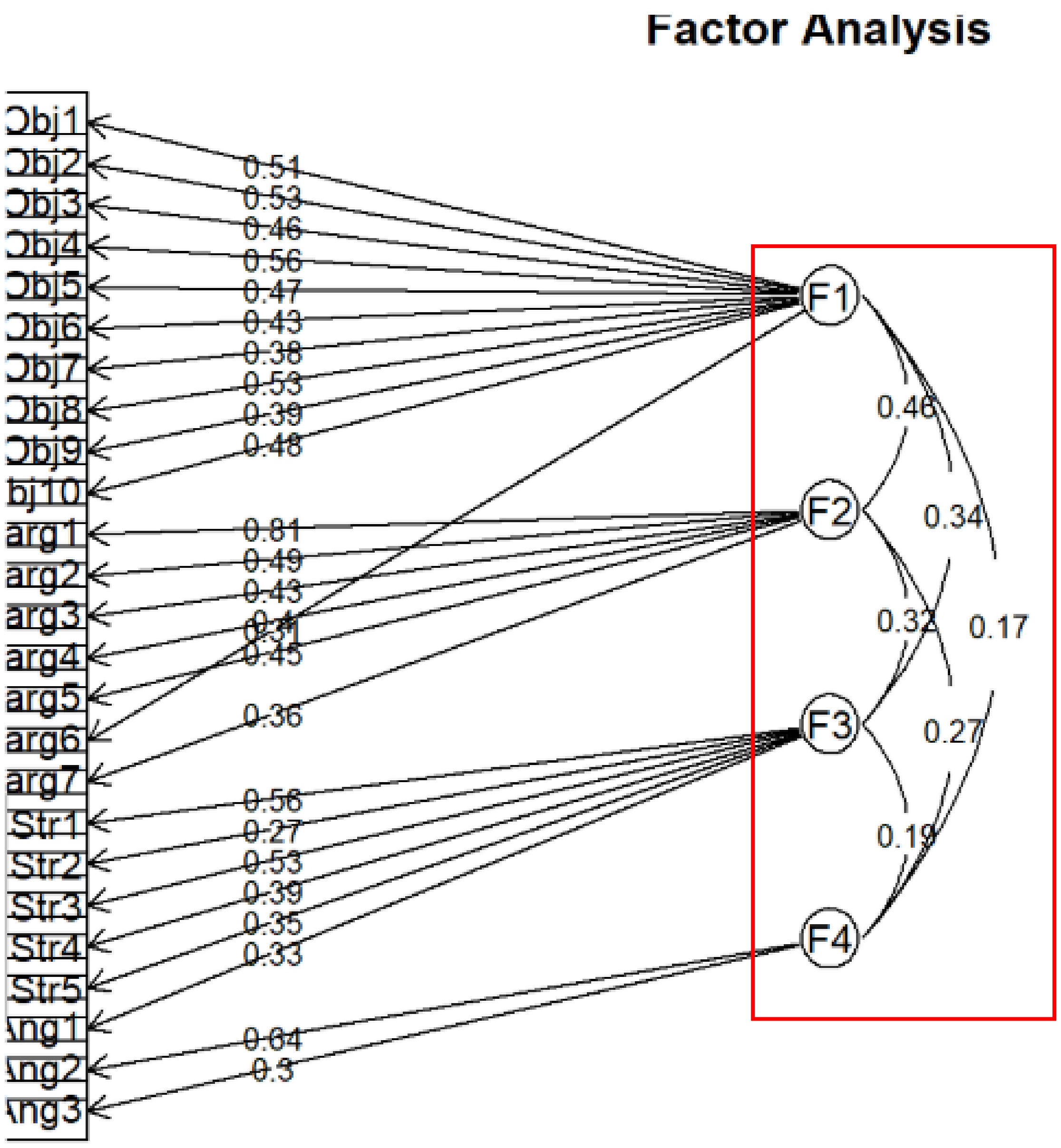
Uniqueness represents the portion of variance that is specific to the variable itself and **not explained** by the common factors.

Summary

Factor	SS Loadings	% of Variance	Cumulative %
1	2.79	11.18	11.18
2	2.03	8.11	19.29
3	1.37	5.47	24.75
4	0.91	3.65	28.41

The SS Loadings column represents the **eigenvalues λ** for each factor

Factor Analysis diagram-Oblimin rotation



- **Factor 1: Assumptions of Beauty and Sexual Objectification** (11 items; Obj1-Obj10 and Marg6)
- **Factor 2: Silenced and Marginalized** (6 items; Marg1-Marg5, Marg7)
- **Factor 3: Strong Woman Stereotype** (6 items; Str1-Str5, Ang1)
- **Factor 4: Angry Woman Stereotype** (2 items; Ang2 and Ang3)

Inter-Factor Correlations

	1	2	3	4
1	—	0.46	0.34	0.17
2		—	0.32	0.27
3			—	0.19
4				—

The four factors are positively correlated with each other, with correlations ranging from 0.17 to 0.46

Model fit

- Chi-square test. If p>0.05, the model **fits the data well** (fail to reject H_0).
- Root mean square error of approximation (**RMSEA**). It is suggested RMSEA < 0.05 (but definitely < 0.10) for a good fit.

Model Fit Measures

	RMSEA 90% CI				Model Test		
RMSEA	Lower	Upper	TLI	BIC	χ^2	df	p
0.00	0.00	0.02	1.03	−955.52	189.19	206	0.794

An RMSEA value close to 0, indicating a good fit between the model and the observed data.

Model fits the data well (p>0.05)

Factor scores

	Ang3	Score Fact...	Score Fact...	Score Fact...	Score Fact...
1	2	-0.871	-1.411	0.346	-0.598
2	2	0.379	-0.942	0.671	0.395
3	2	0.400	0.385	0.124	1.438
4	1	-1.080	-1.230	0.215	-1.847
5	2	-0.228	0.580	-1.064	0.254
6	3	-0.535	0.470	-0.258	1.026
7	1	0.320	0.576	-1.074	-0.665
8	4	-1.631	-0.642	0.626	0.765
9	3	-0.531	-1.458	-0.774	0.760
10	3	0.214	-0.573	-0.792	0.323
11	3	0.504	0.086	0.149	-0.528
12	3	0.557	-0.096	1.151	0.594
13	2	-1.256	-1.134	0.907	0.080
14	2	-0.519	0.983	0.522	0.052
15	2	1.079	-1.206	-0.106	-1.187
16	3	0.927	0.203	-0.218	0.525
17	3	-0.698	-1.464	0.711	1.470
18	4	-0.388	-0.774	0.290	-0.667
19	2	0.753	0.531	1.426	-0.958
20	3	1.511	1.960	1.831	0.874
21	3	-0.655	-0.349	0.247	0.705
22	4	-0.771	-0.007	0.955	1.170
23	3	-0.356	-0.401	0.261	0.816
24	2	1.068	-0.237	-1.636	0.277
25	2	-0.670	-1.088	-0.006	-0.461

For Oblimin rotation:

- **ten Berge's method**
- **Bartlett's method**

Reliability analysis

Cronbach's alpha reflects **internal consistency**

Factor 1

Scale Reliability Statistics

	Cronbach's α
scale	0.78

Factor 2

Scale Reliability Statistics

	Cronbach's α
scale	0.73

Factor 3

Scale Reliability Statistics

	Cronbach's α
scale	0.59

Factor 4

Scale Reliability Statistics

	Cronbach's α
scale	0.40

Note: A factor with **more items** can artificially inflate Cronbach's alpha.