Basic Probabilities, Sampling Distribution

Mu He

Table of contents I

1 Factor Analysis Model

2 Detailed Explanation of L and F

Factor Analysis Model

Definition

The model describes a set of p observations for each of n individuals using k common factors ($f_{i,j}$), where the number of factors is less than the number of observations ($k < p$). Each individual has k common factors, and these factors are related to the observations through a factor loading matrix ($L \in \mathbb{R}^{p \times k}$).

Mathematical Representation

$$x_{i,m} - \mu_i = l_{i,1} f_{1,m} + \cdots + l_{i,k} f_{k,m} + \varepsilon_{i,m}$$

where:

- $x_{i,m}$: Value of the i-th observation of the m-th individual.
- μ_i: Mean of the i-th observation.
- $l_{i,j}$: Loading for the i-th observation of the j-th factor.
- $f_{j,m}$: Value of the j-th factor for the m-th individual.
- $\varepsilon_{i,m}$: Unobserved stochastic error term with mean zero and finite variance.

Matrix Notation

$$X - M = LF + \varepsilon$$

where:

- $X \in \mathbb{R}^{p \times n}$: Observation matrix.
- $L \in \mathbb{R}^{p \times k}$: Loading matrix.
- $F \in \mathbb{R}^{k \times n}$: Factor matrix.
- $\varepsilon \in \mathbb{R}^{p \times n}$: Error term matrix.
- $M \in \mathbb{R}^{p \times n}$: Mean matrix with elements $M_{i,m} = \mu_i$.

Assumptions on F

1. F and ε are independent.
2. $\mathrm{E}(F) = 0$, where E denotes the expectation.
3. $\mathrm{Cov}(F) = I$, ensuring that the factors are uncorrelated, where I is the identity matrix.

Covariance Structure

If $\mathrm{Cov}(X - M) = \Sigma$, then:

$$\Sigma = \mathrm{Cov}(X - M) = \mathrm{Cov}(LF + \varepsilon)$$

Using the properties of covariance and the assumptions on F:

$$\Sigma = L\mathrm{Cov}(F)L^T + \mathrm{Cov}(\varepsilon)$$

Given $\mathrm{Cov}(F) = I$ and setting $\Psi = \mathrm{Cov}(\varepsilon)$:

$$\Sigma = LL^T + \Psi$$

Orthogonal Transformation

For any orthogonal matrix Q:

- If we set $L' = LQ$ and $F' = Q^T F$, the criteria for being factors and factor loadings still hold.
- Therefore, a set of factors and factor loadings is unique only up to an orthogonal transformation.

Detailed Explanation of L and F

Loading Matrix (L)

- The loading matrix L contains the coefficients (loadings) that relate each observed variable to the underlying factors.
- Each element $l_{i,j}$ in L represents the contribution of the j-th factor to the i-th observed variable.
- High loadings indicate a strong relationship between the observed variable and the factor.

Factor Matrix (F)

- The factor matrix F contains the scores for each factor for each individual.
- Each element $f_{j,m}$ in F represents the score of the j-th factor for the m-th individual.
- The factors are assumed to be uncorrelated and have a mean of zero.

Example Data

Let's consider a small dataset of student performance with three observed variables: test scores, attendance, and participation.

	test_scores	attendance	participation
1	85	95	80
2	90	98	85
3	78	85	70
4	92	99	90
5	88	96	88
6	76	80	65
7	95	100	95
8	89	97	85
9	84	93	78
10	91	99	92

Example Data

```
Factor Analysis using method =  minres
Call: fa(r = data, nfactors = 2, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
               MR1  MR2   h2    u2 com
test_scores   0.90 0.43 0.99 0.010 1.4
attendance    0.86 0.44 0.94 0.060 1.5
participation 0.90 0.40 0.97 0.026 1.4

                       MR1  MR2
SS loadings           2.37 0.54
Proportion Var        0.79 0.18
Cumulative Var        0.79 0.97
Proportion Explained  0.82 0.18
Cumulative Proportion 0.82 1.00

Mean item complexity =  1.4
Test of the hypothesis that 2 factors are sufficient.

df null model =  3  with the objective function =  6 with Chi Square =  42.98
df of  the model are -2  and the objective function was  0

The root mean square of the residuals (RMSR) is  0
The df corrected root mean square of the residuals is  NA

The harmonic n.obs is  10 with the empirical chi square  0  with prob <  NA
The total n.obs was  10  with Likelihood Chi Square =  0  with prob <  NA

Tucker Lewis Index of factoring reliability =  1.094
```

Loading Matrix (L)

The loading matrix (L) shows the relationship between the observed variables and the underlying factors.

```
Loadings:
              MR1   MR2
test_scores   0.899 0.426
attendance    0.864 0.440
participation 0.902 0.401

                 MR1   MR2
SS loadings    2.367 0.536
Proportion Var 0.789 0.179
Cumulative Var 0.789 0.968
```

Factor Matrix (F)

The factor matrix (F) contains the scores for each factor for each individual.

```
             MR1         MR2
 [1,] -0.3117150  0.07431425
 [2,]  0.3045455  0.42725116
 [3,] -1.2500270 -0.66730981
 [4,]  0.7283643  0.36928079
 [5,]  0.3614808 -0.08057218
 [6,] -1.6060008 -0.90711113
 [7,]  1.2276378  0.39865091
 [8,]  0.2460520  0.26546776
 [9,] -0.4691955 -0.03907437
[10,]  0.7688578  0.15910262
```