Contents

1 Factor Analysis

Bibliography of the book: Applied Multivariate Statistical Analysis of Johnson and Wicherin

1.1 Introduction

Introduction. Factor analysis aims at measuring the unobservable (aka construct, concept) eg the IQ. The factors manifest by symptoms.

$$X_{p\times 1} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \quad \mu_{p\times 1} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & & \cdots & \sigma_{pp} \end{pmatrix}$$

And we will consider m factors in a vector

$$F_{m\times 1} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix}$$

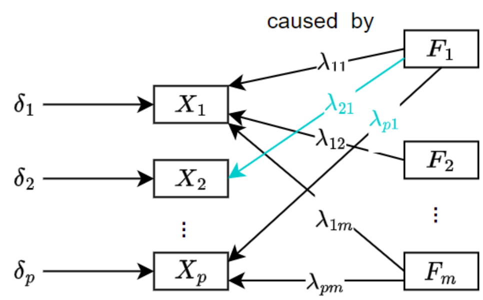

Figure 1: Caption

We also assume that F_i, $i = 1, .., m$ cannot explain perfectly X_1 as such we add an error term δ_1.
Note that those factors are independent of each other that is:

$$\begin{aligned} X_1 &= \mu_1 + \lambda_{11}F_1 + \lambda_{12}F_2 + ..\lambda_{1m}F_m + \delta_1 \\ &\vdots \\ X_j &= \mu_j + \lambda_{j1}F_1 + \lambda_{j2}F_2 + ..\lambda_{jm}F_m + \delta_j \\ &\vdots \\ X_p &= \mu_p + \lambda_{p1}F_1 + \lambda_{p2}F_2 + ..\lambda_{pm}F_m + \delta_p \end{aligned}$$

Or in matrix form:

$$\underset{p\times 1}{\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}} = \underset{p\times 1}{\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}} + \underset{p\times m}{\begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \vdots & & & \vdots \\ \lambda_{j1} & & \ddots & \lambda_{jm} \\ \vdots & & & \vdots \\ \lambda_{p1} & & & \lambda_{pm} \end{pmatrix}} \underset{m\times 1}{\begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix}} + \underset{p\times 1}{\begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{pmatrix}}$$

Definition (Factor Model Equation). All in all:

$$X = \mu + \Lambda F + \delta \iff X - \mu = \Lambda F + \delta$$

Assumptions (Factor Model Equation). We suppose that we know $E(X) = \mu$, $cov(X) = \Sigma$ and we will assume that $E(F) = 0, E(\delta) = 0$. Which if we feed to the FME equation:

$$X-\mu=\Lambda F+\delta \iff \underbrace{\mathbb{E}(X-\mu)}_{=0}=\mathbb{E}(\Lambda F)+\mathbb{E}(\delta)$$

Regarding the covariance (which is centered see on the right):

$$cov(F)=E(FF^T)=I; \quad \text{since } E(F)=0$$

And

$$cov(\delta)=\Psi=diag(\Psi_{11},\Psi_{22},..,\Psi_{pp})$$

Along with

$$cov(F\delta)=0$$

Definition (the Explanatory FM). If all the assumptions hold then the factor model is known as "explanatory factor model" or "orthogonal factor model".

- Oblique FM. We call oblique FM whenever there exists a correlation between the F.

- Families. There are two types of FM:

- Explanatory: we say that $X-\mu=\Lambda F+\delta$ where we explore to find hidden factors.
- Confirmatory. This is when we know clearly the connections between the variables.

- Covariance matrix of the data.

$$\begin{aligned}
\Sigma &= \mathbb{E}\left[(X-\mu)(X-\mu)^T\right]; \quad X-\mu=\Lambda F+\delta \\
&= \mathbb{E}\left[(\Lambda F+\delta)(\Lambda F+\delta)^T\right] \\
&= \mathbb{E}\left[\Lambda F F^T\Lambda^T+\Lambda F\delta^T+\delta F^T\Lambda^T+\delta\delta^T\right] \\
&= \Lambda\mathbb{E}(FF^T)\Lambda^T+\Lambda\mathbb{E}(F\delta^T)+\mathbb{E}(\delta F^T)\Lambda^T+E(\delta\delta^T) \\
&= \Lambda I\Lambda^T+\Lambda\times 0+0\times\Lambda^T+\Psi \\
&= \Lambda I\Lambda^T+\Psi=\Lambda\Lambda^T+\Psi
\end{aligned}$$

If we know Λ and Ψ we know Σ. What is Λ the loading matrix ?

$$\Lambda=\begin{pmatrix}\lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \vdots & & & \\ \lambda_{p1} & & & \lambda_{pm}\end{pmatrix}$$

They are the loading:

$$X_j=\lambda_{j1}F_1+\lambda_{j2}F_2+..+\lambda_{jk}F_k+..+\lambda_{jm}F_m+\delta_j$$

Where λ_{jk} is the loading of the kth factor on the jth X.

Next

$$\begin{aligned}
\Lambda\Lambda^T &= \begin{pmatrix}\lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \vdots & & & \\ \lambda_{p1} & & & \lambda_{pm}\end{pmatrix}\begin{pmatrix}\lambda_{11} & \lambda_{21} & \dots & \lambda_{p1} \\ \vdots & & & \\ \lambda_{1m} & & & \lambda_{pm}\end{pmatrix} \\
&= \begin{pmatrix}\sum_{k=1}^m\lambda_{1k}^2 & \sum_{k=1}^m\lambda_{1k}\lambda_{2k} & \dots & \sum_{k=1}^m\lambda_{1k}\lambda_{pk} \\ \sum_{k=1}^m\lambda_{1k}\lambda_{2k} & \sum_{k=1}^m\lambda_{2k}^2 & \dots & \sum_{k=1}^m\lambda_{2k}\lambda_{pk} \\ \vdots & & & \\ \sum_{k=1}^m\lambda_{1k}\lambda_{pk} & \dots & \dots & \sum_{k=1}^m\lambda_{pk}^2\end{pmatrix}
\end{aligned}$$

Thus:

$$\Sigma = \Lambda\Lambda^T + \Psi = \begin{pmatrix} \sum_{k=1}^m \lambda_{1k}^2 & & \cdots & \\ & \sum_{k=1}^m \lambda_{2k}^2 & \cdots & \cdots \\ \vdots & & & \\ \cdots & \cdots & \cdots & \sum_{k=1}^m \lambda_{pk}^2 \end{pmatrix} + \begin{pmatrix} \Psi_{11} & 0 & \cdots & 0 \\ 0 & \Psi_{22} & \cdots & 0 \\ 0 & & \ddots & 0 \\ 0 & 0 & \cdots & \Psi_{pp} \end{pmatrix}$$

We only compare their variances:

$$\begin{aligned} \sigma_{11} &= \sum_{k=1}^m \lambda_{1k}^2 + \Psi_{11} \\ \sigma_{22} &= \sum_{k=1}^m \lambda_{2k}^2 + \Psi_{22} \\ &\vdots \\ \sigma_{jj} &= \sum_{k=1}^m \lambda_{jk}^2 + \Psi_{jj} \end{aligned}$$

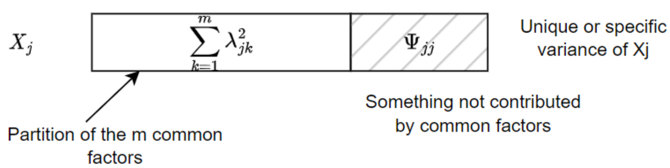

Figure 2: The left part is termed communalities h_i^2

Another concept:

$$\begin{aligned} cov(XF) &= E\left[(X-\mu)(F - \underbrace{E(F)}_{=0})^T\right] \\ &= E\left[(X-\mu)F^T\right], \quad X-\mu = \Lambda F + \delta \\ &= (\Lambda F + \delta)F^T \\ &= \Lambda F F^T + \delta F^T \end{aligned}$$

Then we can pass the expected values on both sides:

$$\begin{aligned} E\left[(X-\mu)F^T\right] &= E(\Lambda F F^T) + E(\delta F^T) \\ &= \Lambda E(FF^T) + 0 \\ &= \Lambda \end{aligned}$$

What we have discussed is a fundamental model, on how the covariance is divided into communalities and specific factors.. Our next topic will be estimation.

- Example.

- We have a sample whereas we only talked about population parameters μ, Σ ...
- Correlation matrix gives clues if we are capable of performing FM
- For PCA we use Bartelet test we test; it is a test on the correlation matrix:

$$\begin{aligned} H_0 &: \ R = I \\ H_1 &: \ R \neq I \end{aligned}$$

1.2 Estimation

Recollect that we derived the formula for Exploratory FA as:

$$X - \mu = \Lambda F + \delta$$

And all in all we have to estimate:

$$\underset{p\times p}{\Sigma} = \underset{(p\times m)(m\times p)}{\Lambda\Lambda^T} + \underset{p\times p}{\Psi} \Rightarrow \text{Estimate } \hat{\Lambda}, \hat{\Psi}$$

- But what is our input(s) ? Solely $\hat{\Sigma} = S_{p\times p}$ since we have collected a dataset $X_{n\times p}$. If n is large and the sampling strategy is appropriate we can write down:

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11} & .. & \hat{\sigma}_{1p} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1} & \dots & \hat{\sigma}_{pp} \end{pmatrix} = S = \begin{pmatrix} s_{11} & .. & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \dots & s_{pp} \end{pmatrix}$$

Where $\hat{\sigma}_{jk} = \hat{s}_{jk}$ for $j = 1, .., p; k = 1, .., p$.

- Estimation. We want to estimate $\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}$ with $tr(\Sigma) = \sum_{j=1}^{p} \sigma_{jj}$, $tr(S) = \sum_{j=1}^{p} s_{jj}$. Further if we perform a Least-square estimation we minimize the errors between the sample and the population:

$$\begin{aligned} G = (S-\Sigma)^2 &= (S-\Sigma)^T(S-\Sigma) \\ &= Tr\left[(S-\Sigma)^T(S-\Sigma)\right], \quad \text{Trace since we only consider variances} \\ &= ||(S-\Sigma)^T|| \\ &= ||S - \Lambda\Lambda^T + \Psi|| \end{aligned}$$

The problem now is that if we calculate the derivative of the G loss we cannot since we do not know Λ and Ψ.

Now to define m the number of factors we extract. We find the degree of freedom (DOF) the number of parameters for Σ unconstrained minus the number of parameters for Σ constrained:

$$\left\{\frac{1}{2}p(p+1)\right\} - \left\{(pm+p) - \frac{1}{2}m(m-1)\right\} = \frac{1}{2}\left[(p-m)^2 - (p+m)\right]$$

Then we have three situations:

1. If DOF $= 0$ then we have a unique solution
2. If $DOF > 0$ we have several solutions
3. If $DOF < 0$ the problem is not determined

As such we have to add constraints such as $\Lambda^T\Psi^{-1}\Lambda = 0$ is diagonal.

- Methods for estimation.

1. Principal component method
2. Principal factor method
3. MLE method

1.2.1 Principal Component Method

For PCA we let the population covariance matrix Σ can be decomposed into its eigenvalues and eigenvectors. If we say that Q_j is the diagonal value and e_j is the eigenvector of size $(p \times 1)$ then by spectral decomposition we have:

$$\Sigma = \sum_{j=1}^{p} Q_j e_j e_j^T$$

Note that in PCA we say λ_j for Q_j since in the FA framework λ_j are the loadings. Next:

$$\begin{aligned}
\Sigma &= Q_1 e_1 e_1^T + Q_2 e_2 e_2^T + .. + Q_p e_p e_p^T \\
&= \begin{pmatrix} \sqrt{Q_1} e_1 & \dots & \sqrt{Q_m} e_m | & \dots & \sqrt{Q_p} e_p \end{pmatrix} \begin{pmatrix} \sqrt{Q_1} e_1 \\ \dots \\ \sqrt{Q_m} e_m \\ - \\ .. \\ \sqrt{Q_p} e_p \end{pmatrix} \\
&= \Lambda \Lambda^T
\end{aligned}$$

- Reducing dimension add the Ψ term. The number of components to be extracted is p and $\Sigma_{p \times p}$ yields the maximum possible component and full assume that the matrix is full rank. Additionally in FA many variables are correlated so we want $m < p$ number of factors. The term in $\{m+1, .., p\}$ are attributed to be error terms. So we want:

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

As such:

$$\Lambda = \begin{pmatrix} \sqrt{Q_1} e_1 & \dots & \sqrt{Q_m} e_m \end{pmatrix}$$

And

$$\Psi = \begin{pmatrix} \sqrt{Q_{m+1}} e_{m+1} & \dots & \sqrt{Q_p} e_p \end{pmatrix} \begin{pmatrix} \sqrt{Q_{m+1}} e_{m+1} \\ \dots \\ \sqrt{Q_p} e_p \end{pmatrix}$$

It means the first factor :

$$\begin{aligned}
(F_1) &: \quad \sqrt{\hat{Q}_1} \hat{e}_1, \quad e_1 = \text{eigenvectors} \\
(F_2) &: \quad \sqrt{\hat{Q}_2} \hat{e}_2 \\
(F_m) &: \quad \sqrt{\hat{Q}_m} \hat{e}_m
\end{aligned}$$

Where the Q are the loadings. On top of that we get from:

$$S = \hat{\Lambda} \hat{\Lambda}^T + \hat{\Psi} \iff \hat{\Psi} = S - \hat{\Lambda} \hat{\Lambda}^T$$

To sum up we have a five steps process:

1. Find S
2. Find the eigenvalues / vectors of S
3. Select m
4. Obtain $\hat{\Lambda}$
5. Obtain $\hat{\Psi}$

Example. From the data matrix X we build its empirical covariance matrix:

$$S = \begin{pmatrix} 100 & 50 & -30 \\ 50 & 64 & -20 \\ -30 & -20 & 49 \end{pmatrix}_{3\times 3}$$

(i) Covariance matrix decomposition. First recollect that we have derived the covariance matrix using two equations:

$$S = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right]; \quad X - \mu = \Lambda F + \delta \Rightarrow \boxed{S = \hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}}$$

Additionally if we let $m = 1 < p$ factor model we end up with a 1D factor model:

$$\hat{\Lambda}_{p\times m} = \hat{\Lambda}_{3\times 1} = \begin{pmatrix} | \\ \Lambda_1 \\ | \end{pmatrix} = \begin{pmatrix} \hat{\lambda}_{11} \\ \hat{\lambda}_{21} \\ \hat{\lambda}_{31} \end{pmatrix}, \quad \hat{\Psi} = \begin{pmatrix} \hat{\Psi}_{11} & 0 & 0 \\ 0 & \hat{\Psi}_{22} & 0 \\ 0 & 0 & \hat{\Psi}_{33} \end{pmatrix}$$

(ii) Eigendecomposition of S. We have

$$det|S - QI| = 0 \iff det|S - \lambda I| = 0$$

We get the characteristic equation which yields eigenvalues $\hat{Q}_i$: $\hat{Q}_1 = 148.23, \hat{Q}_2 = 36.0, \hat{Q}_3 = 28.70$. Additionally we solve for the eigenvectors:

$$\hat{e}_i: \quad (S - Q_i I)e_i = 0; \quad s.t. \;\; e_i^T e_i = 1; \quad i = 1,.,3$$

And eigenvectors $\hat{e}_i$:

$$\hat{e}_1 = \begin{pmatrix} -0.77 \\ -0.54 \\ 0.32 \end{pmatrix} \quad \hat{e}_2 = \begin{pmatrix} 0.21 \\ 0.29 \\ 0.93 \end{pmatrix} \quad \hat{e}_3 = \begin{pmatrix} 0.60 \\ -0.79 \\ 0.11 \end{pmatrix}$$

(iii) Dimensionality reduction. Recollect that we decomposed S for $S = \hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}$ with Ψ that collects all the vectors after $> m$

$$\Sigma = Q_1 e_1 e_1^T + Q_2 e_2 e_2^T + .. + Q_p e_p e_p^T = \left(\sqrt{Q_1}e_1 \quad \dots \quad \sqrt{Q_m}e_m | \quad \dots \quad \sqrt{Q_p}e_p\right) \begin{pmatrix} \sqrt{Q_1}e_1 \\ \dots \\ \sqrt{Q_m}e_m \\ - \\ .. \\ \sqrt{Q_p}e_p \end{pmatrix}$$

Then

$$\hat{\Lambda}_1 = \sqrt{\hat{Q}_1}\hat{e}_1 = \sqrt{148.28}\begin{pmatrix} -0.77 \\ -0.54 \\ 0.32 \end{pmatrix} = \begin{pmatrix} -9.37 \\ -6.57 \\ 4.75 \end{pmatrix}$$

Obtain:

$$\hat{\Psi} \Rightarrow \hat{\Lambda}\hat{\Lambda}^T = \begin{pmatrix} -9.57 \\ 6.57 \\ 4.75 \end{pmatrix}\begin{pmatrix} -9.57 & 6.57 & 4.75 \end{pmatrix} = \begin{pmatrix} 87.80 & 61.56 & -44.51 \\ 61.56 & 43.10 & -31.21 \\ -44.51 & -31.21 & 22.56 \end{pmatrix}$$

$$\hat{\Psi} = S - \hat{\Lambda}\hat{\Lambda}^T = \begin{pmatrix} 12.20 & -11.56 & 14.51 \\ -19.50 & 20.84 & 11.21 \\ 14.51 & 11.21 & 26.44 \end{pmatrix}$$

(iv) Conclusion and caveats. The problem is that the out of diagonal elements are different than zero. Since theoretically:

$$cov(\delta) = \Psi = diag(\Psi_{11}, \Psi_{22}, .., \Psi_{pp})$$

1.2.2 Principal Factor Method

We get a value x_j

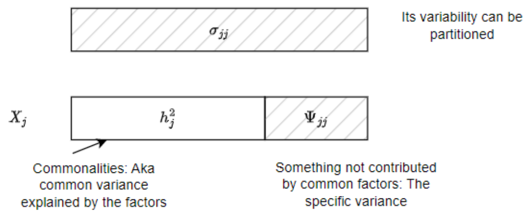

Figure 3: Caption

We can write down:

$$\Sigma = \Lambda\Lambda^T + \Psi; \quad S = \Lambda\Lambda^T + \Psi; \quad \Lambda\Lambda^T = S - \Psi$$

Where recall that in contrast for for PCA we calculated the e.v. directly with the covariance matrix S. Now if we do more manipulation here:

$$S = \begin{pmatrix} s_{11} & \dots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \dots & s_{pp} \end{pmatrix}; \quad \Psi = \begin{pmatrix} \Psi_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Psi_{pp} \end{pmatrix}$$

As a consequence their difference yields:

$$S - \Psi = \begin{pmatrix} s_{11} - \Psi_{11} & \dots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \dots & s_{pp} - \Psi_{pp} \end{pmatrix} = \begin{pmatrix} h_1^2 & \dots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \dots & h_p^2 \end{pmatrix}$$

With $\sigma_{jj} = h_j^2 + \Psi_{jj}$ we take its estimate $s_{jj} = h_j^2 + \Psi_{jj}$ implies $s_{jj} - \Psi_{jj} = h_j^2$, which yield the matrix on very right. We can obtain its eigenvalues with the difference matrix:

$$Q_j^* e_j*, \quad i = 1, .., p$$

We get the same thing as PCA with spectral decomposition of S that is:

$$\hat{\Lambda}^* = \left(\sqrt{Q^*}_1 \hat{e}_1^* \quad \sqrt{Q^*}_2 \hat{e}_2^* \quad \dots \quad \sqrt{Q^*}_m \hat{e}_m^*\right)$$

Here the difference is that we factorize with $S - \Psi$ where we don't know the value of Ψ. That is why we have to initialize Ψ by $\hat{\Psi}_0$ as a consequence this is an iterated procedure.

1.2.3 Maximum Likelihood Estimation

In this case we let $X_i \sim MVN(\mu, \Sigma) = N_p(\mu, \Lambda\Lambda^T + \Psi), i = 1, .., p$. And we write the likelihood as :

$$L(x|\mu, \Sigma) = -\frac{n}{2} log|2\pi\Sigma| - \frac{n}{2} tr(\Sigma^{-1} S) - \frac{n}{2}(\bar{X} - \mu)\Sigma^{-1}(\bar{X} - \mu)^T$$

When we estimate $\hat{\mu} = \bar{X}$ we obtain the likelihood:

$$L(x|\hat{\mu}, \Sigma) = -\frac{n}{2} log|2\pi\Sigma| - \frac{n}{2} Tr(\Sigma^{-1} S)$$

If $\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}$ then

$$L(x|\hat{\mu}, \hat{\Sigma}) = -\frac{n}{2}\left\{log(2\pi(\hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}) + tr(\hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi})^{-1} S\right\}$$

- Lawley (1940). The author has shown that given normal equation we have :

$$S\hat{\Psi}^{-1}\hat{\Lambda} = \hat{\Lambda}(I + \hat{\Lambda}^T\hat{\Psi}^{-1}\hat{\Lambda})$$
$$\hat{\Psi} = diag(S - \hat{\Lambda}\hat{\Lambda}^T)$$

Then the problem is solved by an iterative procedure.

Example. The first step is to look at the correlation matrix, if enough factors are significantly correlated we can go for Factor Analysis. Then here for a 2-factors PCA:

1.3 Model Adequacy Tests

1.3.1 Barlet test

Prerequisite. We say if R is a correlation matrix that have $r_{ij} \geq 0.30$ we can go for FA.

Bartlett's test takes the determinant of R.

$$\chi^2_{p(p-1)/2} = -\left[(n-1) - \frac{(2p+5)}{6^8}\right] \ln |R|$$

Large sample likelihood ratio test:

- Step 1.

$$H_0 : \Sigma = \Lambda\Lambda^T + \Psi$$
$$H_1 : \Sigma = \text{any other pd matrix}$$

- Step 2. We can build a test statistic:

$$-2ln|\Lambda| = 2ln\left|\frac{|\hat{\Sigma}|}{|S_n|}\right| = T_n$$

Where

$$S_n = \frac{n-1}{n}S, \quad \text{for } n \to +\infty, S_n = S$$

- Step 3. Sampling distribution of the test statistic:

$$M \cdot ln\left[\frac{|\hat{\Sigma}|}{|S_n|}\right] \sim \chi^2_{\frac{1}{2}((p-m)^2-(p+m)}$$

The decision is reject H_0 if $D \geq \eta^2_{\frac{1}{2}((p-m)^2-(p+m)}$

1.3.2 Number of factor to be written

The percentage of cumulative variance explain. Eigenvalue criteria. Scree plot.

Eg. let

	F_1	..	F_3
X_1	λ_{11}	..	λ_{1m}
X_2	λ_{21}		
..			
X_p	λ_{p1}		λ_{1pm}
	$\sum_{j=1}^{p} \lambda_{j1}^2$ (%)	..	$\sum_{j=1}^{p} \lambda_{jm}^2$ (%)

Then cumulative %. We can then set a criteria like cumulative % is $\geq 90\%$.

Scree plot.

Factor analysis has two purposes:

- Dimmension reduction
- Interpretability (provide name to each of the factors)

1.3.3 Rotations

If we look at the loadings
- Factor rotation. We can rotate so that the loading as max for the factors (F_1 with $x_1, .., x_3$) 2 min wrt the rest.

Example.

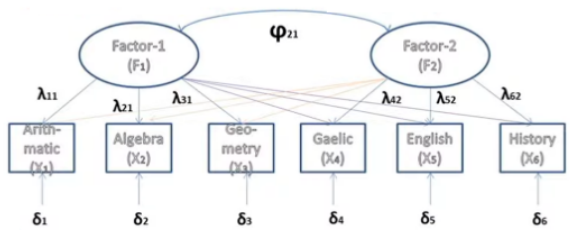

Figure 4: Caption

Gaellic equally loaded with F_1 and F_2 (0.553 and 0.429). English sort of similar. None of the loadings are zero so we need the two factors cannot create interpretable.

	X1	X2	X3	X4	X5	X6
X1	1.00	0.44	0.41	0.29	0.33	0.25
X2		1.00	0.35	0.35	0.32	0.33
X3			1.00	0.16	0.19	0.18
X4				1.00	0.60	0.47
X5					1.00	0.46
X6						1.00

	Variables	Factor-1 F_1	Factor-2 F_2	Communa-lities h_i
$x_1 =$	Gaelic	$\lambda_{11} = 0.553$	$\lambda_{21} = 0.429$	$0.49 = h_1^2 = \lambda_{11}^2 + \lambda_{12}^2$
$x_2 =$	English	$\lambda_{21} = 0.568$	$\lambda_{22} = 0.288$	0.41
	History	$\lambda_{31} = 0.392$	$\lambda_{23} = 0.450$	0.36
...	Arith-matic	$\lambda_{41} = 0.740$	$\lambda_{24} = -0.273$	0.63
	Algebra	$\lambda_{51} = 0.724$	$\lambda_{25} = -0.211$	0.59
$x_6 =$	Geometry	$\lambda_{61} = 0.595$	$\lambda_{26} = -0.132$	0.37
	% variance explained	0.37	0.10	

<u>Eigenvalue of Factor-1:</u> $Q_1 = \sum_{j=1}^{p} \lambda_{j1}^2 = 2.21$

Figure 5: Caption

Plot of loadings

Un-rotated

Variables	Factor-1	Factor-2
Gaelic	0.553	0.429
English	0.568	0.288
History	0.392	0.450
Arith-matic	0.740	-0.273
Algebra	0.724	-0.211
Geometry	0.595	-0.132

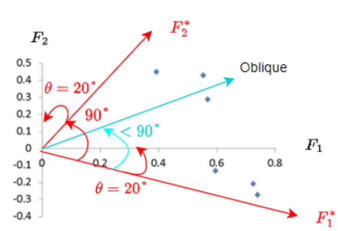

Figure 6: Caption

Rotation: If I know θ then we find F_1^*, F_2^* from F_1 and F_2 Where recall that in PCA we gave:

$$T = \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} \iff \begin{cases} F_1^* & = F_1 cos(\theta) - F_2 sin(\theta) \\ F_2^* & = F_1 sin(\theta) + F_2 cos(\theta) \end{cases}$$

Where $TT^T = T^T T = TT^{-1} = I$. We say it is an orthogonal transformation.

Rotated

Variables	Factor-1	Factor-2
Gaelic	0.369	0.594
English	0.433	0.467
History	0.211	0.558
Arith-matic	0.789	0.001
Algebra	0.752	0.054
Geometry	0.604	0.083

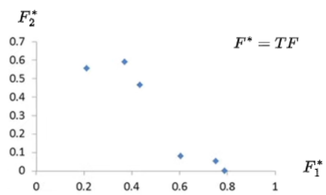

Figure 7: Caption

Almost 0 for the last three variables. First three variables have factor 2 higher for 3 first.

New issue: if we do not orthogonal rotation. Then we get oblique rotation that is the angle is less than 90 degrees. So there is correlation between the two:

$$\Sigma = \Lambda\Lambda^T + \Psi$$
$$X - \mu = \Lambda F + \delta$$

In the oblique/orthogonal setting the model changed to be using $F^* = TF$

$$X - \mu = \Lambda F^* + \delta$$
$$\Sigma = \Lambda \underbrace{TT^T}_{=I} \Lambda^T + \Psi = \Lambda\Lambda^T + \Psi$$

So it will keep the same amount of information from the variance standpoint. But from interpretability standpoint it improves.

Possible transformations.
- Orthogonal. eg varimax rotation, equimax, ..

- Oblique. Now when you know m, the test model.. Then FM can be used for something else. Eg If I go for orthogonal factor model so $F_1 \perp F_i$ we use them for regression analysis. We have factor scores:

$$\begin{array}{l|cccc|cccc|l}
\text{i} & X_1 & X_2 & .. & X_p & F_1 & F_2 & ... & F_m & \\
\text{height1} & x_{11} & \dots & \dots & x_{1p} & f_{11} & f_{12} & \dots & f_{1m} & \\
2 & x_{21} & \ddots & & \vdots & \vdots & \ddots & & \vdots & \\
.. & \vdots & & & \vdots & \vdots & & \ddots & \vdots & \\
\text{i} & \vdots & & \ddots & \vdots & f_{i1} & f_{i2} & \dots & f_{im} & \\
.. & \vdots & & & \vdots & \vdots & & \ddots & \vdots & \\
\text{n} & x_{n1} & \dots & \dots & x_{np} & \vdots & & & & \\
\hline
 & & .. & & & & & & & n \times m
\end{array}$$

Method available

$$X - \mu = \Lambda F + \delta$$

It is different than regression where $y = X\beta + \varepsilon$ rather we use $\hat{f}_c$ ie weighted LS, regression method, MLE; see Johnson and wicher book.

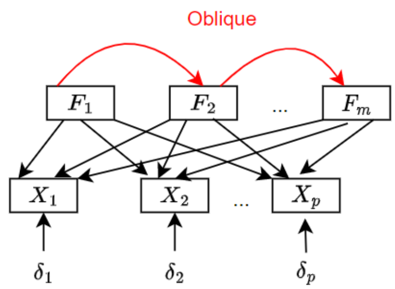

Figure 8: Caption

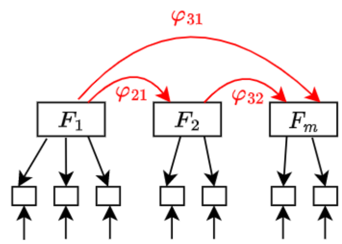

Figure 9: Caption

$$\begin{aligned} X_1 - \mu_1 &= \lambda_{11} F_1 + \delta_1 \\ X_2 - \mu_2 &= \lambda_{21} F_1 + \delta_2 \\ X_3 - \mu_3 &= \lambda_{31} F_1 + \delta_3 \\ X_4 - \mu_4 &= \lambda_{42} F_2 + \delta_4 \end{aligned} \iff \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix}$$

Equivalently:

$$X - \mu = \Lambda F + \delta$$

And

$$cov(X) = \Sigma = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right] = \Lambda \Phi \Lambda^T + \Phi$$

Since we do oblique factor. For confirmatory we have:

$$\Lambda I \Lambda^T +$$

And $E(FF^T) \neq I = \varphi$

But if

- EFA: $\Lambda\Lambda^T + \Psi$
- CFA: $\Lambda\Lambda^T + \Psi$