- regression method (conditional normal distribution) — LNp.3-16.
  - Suppose that the common factors ($\hat F$) and the specific factors ($\hat\varepsilon$) are jointly normally distributed

LNp.5-12

- $\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$ has an $N_p(\mathbf{0}, \mathbf{LL}' + \boldsymbol{\Psi})$ distribution
- Moreover, the joint distribution of $(\mathbf{X} - \boldsymbol{\mu})$ and $\mathbf{F}$ is $N_{m+p}(\mathbf{0}, \boldsymbol{\Sigma}^*)$, where

$\begin{bmatrix} X-\mu \\ F \end{bmatrix}$

$$\underset{(m+p)\times(m+p)}{\boldsymbol{\Sigma}^*} = \left[\begin{array}{c|c} \boldsymbol{\Sigma} = \underset{(p\times p)}{\mathbf{LL}'} + \boldsymbol{\Psi} & \underset{(p\times m)}{\mathbf{L}} \\ \hline \underset{(m\times p)}{\mathbf{L}'} & \underset{(m\times m)}{\mathbf{I}} \end{array}\right]$$

$cov(X-\mu, F)$
$= cov(LF+\varepsilon, F)$
$= L\, cov(F,F) + cov(\varepsilon, F)$ $\quad 0$

- the conditional distribution of $\mathbf{F}|\mathbf{x}$ is multivariate normal with ($\because$ factor model)

LNp.3-4 result3-14

$$\begin{cases} \text{mean} = E(\mathbf{F}|\mathbf{x}) = \underbrace{}_{E(F)+}\mathbf{L}'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ \text{covariance} = \text{Cov}(\mathbf{F}|\mathbf{x}) = \mathbf{I} - \mathbf{L}'\boldsymbol{\Sigma}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}\mathbf{L} \end{cases}$$

similar to the mean structure of response in regression analysis.

$\because$ factor model

- the jth factor score vector is given by

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) = \boxed{\hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}}(\mathbf{x}_j - \bar{\mathbf{x}})$$

$\hookrightarrow$ jth row of X.

- Denote the scores generated by the weighted least squares by $\hat{\mathbf{f}}_j^{LS}$ and those by the regression method by $\hat{\mathbf{f}}_j^R$. Because $\hat f_j^R = (I + \hat L'\hat\Psi^{-1}\hat L)^{-1}\hat L'\hat\Psi^{-1}(x_j - \bar x)$

$$\underset{(m\times p)}{\hat{\mathbf{L}}'}(\hat{\mathbf{L}}\underset{(p\times p)}{\hat{\mathbf{L}}'} + \hat{\boldsymbol{\Psi}})^{-1} = (\mathbf{I} + \underset{(m\times m)}{\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}})^{-1}\underset{(m\times p)}{\hat{\mathbf{L}}'}\underset{(p\times p)}{\hat{\boldsymbol{\Psi}}^{-1}}$$ (exercise 9.6, textbook)

$$\hat{\mathbf{f}}_j^{LS} = (\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}(\mathbf{I} + \hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})\hat{\mathbf{f}}_j^R = (\mathbf{I} + (\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1})\hat{\mathbf{f}}_j^R = (I + \hat\Delta^{-1})\hat f_j^R$$

For maximum likelihood estimates $(\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1} = \hat{\boldsymbol{\Delta}}^{-1}$ and if the elements of this diagonal matrix are close to zero, the regression and generalized least squares methods will give nearly the same factor scores.

- In an attempt to reduce the effect of a (possible) incorrect determination of the number of factors, $\mathbf{S}$ is often used for $\hat{\boldsymbol{\Sigma}}$, rather than $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$, i.e,

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$$

may not be a good approximation.

or, if a correlation matrix is factored,

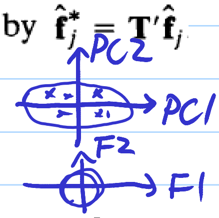
$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}_z'\mathbf{R}^{-1}\mathbf{z}_j$$

$\hat f_j = (\hat P \Lambda^{1/2})'(\hat P \Lambda^{-1}\hat P')(x_j - \bar x)$

where $\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \bar{\mathbf{x}})$

- If rotated loadings $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ are used in place of the original loadings the subsequent factor scores, $\hat{\mathbf{f}}_j^*$, are related to $\hat{\mathbf{f}}_j$ by $\hat{\mathbf{f}}_j^* = \mathbf{T}'\hat{\mathbf{f}}_j$

$\Lambda^{1/2}\underbrace{\hat P'\hat P}_{= I}\Lambda^{-1}\hat P'$ $= \Lambda^{-1/2}\hat P' = \begin{bmatrix} \hat e_1' \\ \hat e_2' \\ \vdots \end{bmatrix}$



- A strategy for factor analysis
  1. perform a principal component factor analysis
  2. perform a maximum likelihood factor analysis
  3. compare the solutions obtained from the 2 factor analyses
  4. repeat the 1st 3 steps for other number of common factors m
  5. for large data sets, split them in half and perform a FA on each part

❖ **Reading**: Textbook, 9.4, 9.5, 9.6