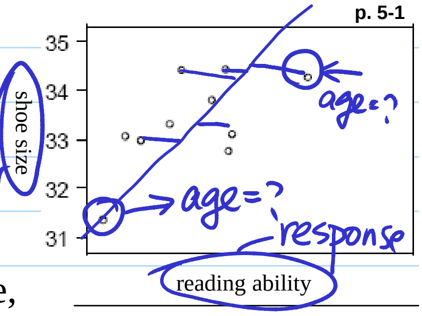
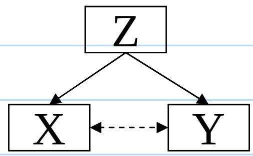
Factor Analysis

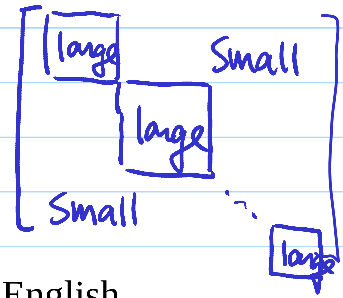


- A motivating example: for children in elementary school
  - observed variables: shoe size and reading ability
    - there exists strong correlation between them
  - latent (lurking) variable: age
  - **Q**: can we extract information about the latent variable, called *factor*, from the observed variables? If yes, what information? not the "factor" in DOE.
- Purpose of factor analysis



cor matrix

  - reduce high dimensional data down to just a few representative variables (similar to PCA)
  - describe the relationship between many variables (as captured by the covariance/correlation matrix) by a few underlying, but unobservable (latent) variables



    - *Suppose variables can be grouped by their correlation*. Then, it is conceivable that each group of variables represents a single underlying factor. For example,
      - 1st group of variables: test scores in classics, French, English, mathematics, and music ⇒ suggest an underlying "intelligence" factor
      - 2nd group of variables representing physical-fitness scores ⇒ might correspond to another factor

- Modeling (orthogonal factor model)
  - The observable random vector $\mathbf{X}$, with p components, has mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
  - $\mathbf{X}$ is linearly dependent upon a few unobservable random variables $F_1, F_2, \ldots, F_m$, called common factors, and p additional sources of variation $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$, called *errors* or, sometimes, *specific factors*.

c.f. in regression, $y=\beta_0+\beta_1 x+\beta_1 x^2+\varepsilon$

strong assumption ①

$$
\begin{cases}
X_1-\mu_1 = \ell_{11}F_1+\ell_{12}F_2+\cdots+\ell_{1m}F_m+\varepsilon_1 \\
X_2-\mu_2 = \ell_{21}F_1+\ell_{22}F_2+\cdots+\ell_{2m}F_m+\varepsilon_2 \\
\vdots \qquad\qquad\qquad \vdots \\
X_p-\mu_p = \ell_{p1}F_1+\ell_{p2}F_2+\cdots+\ell_{pm}F_m+\varepsilon_p
\end{cases}
$$

c.f. regression model (linear model)

or, in matrix notation,

$$Y = X\beta + \varepsilon$$

$$\underset{(p\times 1)}{\mathbf{X}-\boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}}\ \underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$$

factor

    - The coefficient ℓ_{ij} is called the *loading* of the ith variable on the jth factor
    - Note: $m < p$. (otherwise, why bother?) We want m as small as possible.
    - Note: $F_1, F_2, \ldots, F_m, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ are unobservable. This distinguishes the factor model from the regression model
  - assume that

assumption 2

$$E(\mathbf{F}) = \underset{(m\times 1)}{\mathbf{0}}, \qquad \mathrm{Cov}(\mathbf{F}) = E[\mathbf{F}\mathbf{F}'] = \underset{(m\times m)}{\mathbf{I}}$$

- assume that

assumption 3

$$E(\boldsymbol{\varepsilon}) = \underset{(p\times 1)}{\mathbf{0}}, \qquad \text{Cov}(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \underset{(p\times p)}{\boldsymbol{\Psi}} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

- assume that $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent, so

assumption 4.

$$\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \underset{(p\times m)}{\mathbf{0}}$$

- Some results of the orthogonal factor model

- $$(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' = (\mathbf{LF}+\boldsymbol{\varepsilon})(\mathbf{LF}+\boldsymbol{\varepsilon})'$$
$$= (\mathbf{LF}+\boldsymbol{\varepsilon})((\mathbf{LF})'+\boldsymbol{\varepsilon}')$$
$$= \mathbf{LF}(\mathbf{LF})' + \boldsymbol{\varepsilon}(\mathbf{LF})' + \mathbf{LF}\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$$

$$\begin{bmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{pp} \end{bmatrix} =$$

so that

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'$$
$$= \mathbf{L}\underbrace{E(\mathbf{FF}')}_{I}\mathbf{L}' + \underbrace{E(\boldsymbol{\varepsilon}\mathbf{F}')}_{=0}\mathbf{L}' + \mathbf{L}\underbrace{E(\mathbf{F}\boldsymbol{\varepsilon}')}_{=0} + \underbrace{E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')}_{=\Psi}$$
$$= \mathbf{LL}' + \boldsymbol{\Psi}$$

assumption.

$(LL')_{ii}$ =

- $$\underbrace{\sigma_{ii}}_{\text{Var}(X_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2}_{\text{communality}} + \underbrace{\psi_i}_{\text{specific variance}} = h_i^2 + \psi_i$$

  - $h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2$ is called the ith *communality*, which represents the variance shared with the other observed variables via the common factor
  - ψ_i is called the ith *specific variance*, which relates to the variability in X_i not shared with other variables