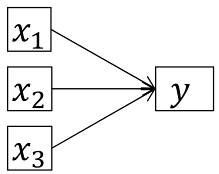
# 3 Structural Equation Modeling

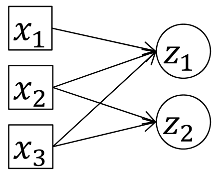
# 構造方程式モデリング/共分散構造解析

## 3.1 Objectives

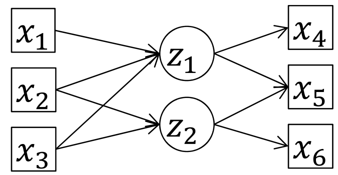
- Multiple regression analysis, PCA and many of other methods deal with only limited model structures.




Multiple regression analysis



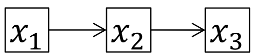

PCA

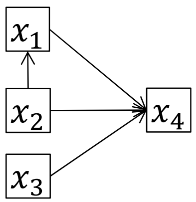


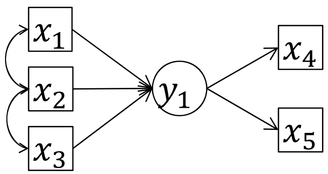

Canonical correlation analysis

## 3.1 Objectives

- Structural equation modeling (SEM) flexibly accepts various types of models that conform to real data. For example,




- These structures include those in the previous page. Hence, SEM can also handles the models of multiple regression analysis, PCA, and other analyses.

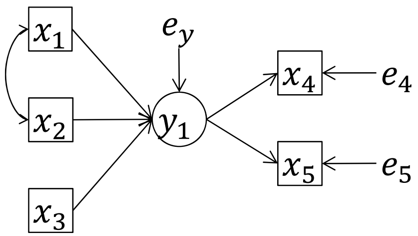
## 3.1 Objectives

- SEM is a general form of some multivariate analysis techniques.
- SEM can flexibly address most types of linear structures.
- SEM is hypothesis-driven. It does not tell us the best structure to explain the data. It tells us whether the data statistically fit into the model or not.

# 3.2 Pass diagram

- Graphically express the relationships between variables. Learn some general rules.

$x_1 \perp x_3$
$x_2 \perp x_3$




$\boxed{x_i}$ Observed variable (square) $e_i \rightarrow$ Random error

$y_i$ Latent variable (circle) $\longleftrightarrow$ Correlation

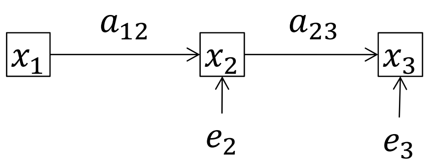
# 3.3 Covariance structure (共分散構造)

- The covariance structure is a covariance matrix, each element of which is expressed in the form formula.
- This matrix includes the information of the model structure.
- Let's see a few examples.
- For 3-variable problem, the covariance matrix is given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{Var}[x_1] & \mathrm{Cov}[x_1, x_2] & \mathrm{Cov}[x_1, x_3] \\ & \mathrm{Var}[x_2] & \mathrm{Cov}[x_2, x_3] \\ & & \mathrm{Var}[x_3] \end{bmatrix}$$

- We derive all elements of this matrix.

# 3.3.1 Link model




$x_i$: Observed variables. $\mathrm{E}[x_i] = 0$, $\mathrm{Var}[x_i] = 1$. Gaussian, Ideally.

$e_i$: Unobservable error. $e_i \sim N(0, \sigma_i^2)$, $e_i \perp e_j$

$a_{ij}$: Influence from one variable to another.

Structural equations

$$\begin{cases} x_1 = x_1 \\ x_2 = a_{12} x_1 + e_2 \\ x_3 = a_{23} x_2 + e_3 \end{cases} \qquad (3.1)$$

# Mathematical review V: Variance

$$\begin{aligned} \mathrm{Var}[x] &= \frac{1}{n} \sum (x_i - \bar{x})^2 \\ &= \frac{1}{n} \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{1}{n} \sum (x_i^2) - \frac{2}{n} \bar{x} \sum (x_i) + \frac{1}{n} \sum (\bar{x}^2) \\ &= \mathrm{E}[x^2] - 2\bar{x}\mathrm{E}[x] + \mathrm{E}[x]^2 \\ &= \mathrm{E}[x^2] - 2\mathrm{E}[x]^2 + \mathrm{E}[x]^2 \\ &= \mathrm{E}[x^2] - \mathrm{E}[x]^2 \end{aligned}$$

# 3.3.1 Link model

Variance of each variables

$$\mathrm{Var}[x_1] = \sigma_1^2 \quad (3.2)$$

$$\mathrm{Var}[x_2] = \mathrm{Var}[a_{12}x_1 + e_2] \quad (3.3)$$

$$= \mathrm{E}[(a_{12}x_1 + e_2)^2] - \mathrm{E}[a_{12}x_1 + e_2]^2$$

1st term

$$\mathrm{E}[(a_{12}x_1 + e_2)^2] = \mathrm{E}[a_{12}^2x_1^2 + 2a_{12}x_1e_2 + e_2^2]$$

$$= \mathrm{E}[a_{12}^2x_1^2] + \mathrm{E}[2a_{12}x_1e_2] + \mathrm{E}[e_2^2]$$

$\mathrm{Var}[e_2] \equiv \sigma_2^2 \Rightarrow$

$$= a_{12}^2\mathrm{E}[x_1^2] + 2a_{12}\mathrm{E}[x_1e_2] + \sigma_2^2$$

$$= a_{12}^2\sigma_1^2 + 2a_{12}\mathrm{E}[x_1e_2] + \sigma_2^2$$

$$= a_{12}^2\sigma_1^2 + \sigma_2^2$$

# 3.3.1 Link model

Note 1:

$$\begin{cases} \mathrm{Var}[x_1] = \mathrm{E}[x_1^2] - \mathrm{E}[x_1]^2 \\ \mathrm{E}[x_1] = 0 \end{cases}$$

$$\Rightarrow \mathrm{Var}[x_1] = \mathrm{E}[x_1^2] = \sigma_1^2$$

Note 2: Moment (積率) of two independent random variables (確率変数)

When $x \perp y$,

$$E[xy] = \sum x_k y_k = \boldsymbol{x}^T\boldsymbol{y} = 0$$

# 3.3.1 Link model

2nd term

$$\mathrm{E}[a_{12}x_1 + e_2]^2 = (a_{12}\mathrm{E}[x_1] + \mathrm{E}[e_2])^2$$

$$= (0 + 0)^2$$

$$= 0$$

$$\mathrm{Var}[x_3] = \mathrm{Var}[a_{23}x_2 + e_3] \quad (3.4)$$

$$= \mathrm{Var}[a_{23}(a_{12}x_1 + e_2) + e_3]$$

$\mathrm{Var}[e_3] \equiv \sigma_3^2$

$$= a_{12}^2a_{23}^2\sigma_1^2 + a_{23}^2\sigma_2^2 + \sigma_3^2$$

# 3.3.1 Link model

Covariance between variables

$$\mathrm{Cov}[x_1, x_2] = \frac{\sum(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{n\sqrt{\mathrm{Var}[x_1]\mathrm{Var}[x_2]}} \quad (3.5)$$

$$= \frac{\sum(x_{1i})(x_{2i})}{n}$$

$$= \mathrm{E}[x_1x_2]$$

$$= \mathrm{E}[x_1(a_{12}x_1 + e_2)]$$

$$= a_{12}\mathrm{E}[x_1^2] + \mathrm{E}[x_1e_2]$$

$$= a_{12}\sigma_1^2$$

# 3.3.1 Link model

Covariance between variables

$$\begin{aligned}
\mathrm{Cov}[x_1, x_3] &= \mathrm{E}[x_1 x_3] \\
&= \mathrm{E}[x_1(a_{23}x_2 + e_3)] \\
&= \mathrm{E}[x_1(a_{23}(a_{12}x_1 + e_2) + e_3)] \\
&= \mathrm{E}[a_{12}a_{23}x_1^2 + a_{23}x_1 e_2 + x_1 e_3] \\
&= \mathrm{E}\left[a_{12}a_{23}x_1^2\right] + 0 + 0 \\
&= a_{12}a_{23}\sigma_1^2
\end{aligned} \tag{3.6}$$

# 3.3.1 Link model

Covariance between variables

$$\mathrm{Cov}[x_2, x_3] = \mathrm{E}[x_2 x_3] \tag{3.7}$$

$$= a_{23}(a_{12}^2\sigma_1^2 + \sigma_2^2)$$

# 3.3.1 Link model

- Covariance structure of the 3-link model is then

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & a_{12}\sigma_1^2 & a_{12}a_{23}\sigma_1^2 \\ & a_{12}^2\sigma_1^2 + \sigma_2^2 & a_{23}(a_{12}^2\sigma_1^2 + \sigma_2^2) \\ & & a_{12}^2a_{23}^2\sigma_1^2 + a_{23}^2\sigma_2^2 + \sigma_3^2 \end{bmatrix} \tag{3.8}$$

- Although $\boldsymbol{\Sigma}$ is called the covariance structure, because the variables are normalized, the matrix is actually the correlation matrix.

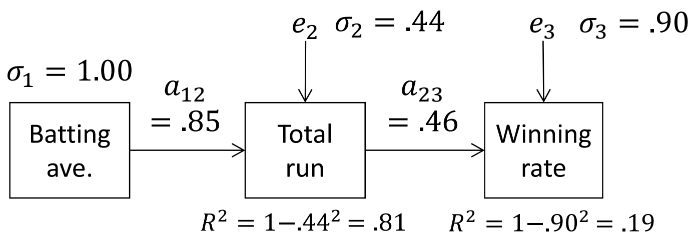
# 3.3.1 Link model

- The correlation matrix ($\boldsymbol{C}$)of 3 baseball parameters

| | Batting ave. | Total run | Winning rate |
|---|---|---|---|
| Batting ave. | 1.00 | .81 | .49 |
| Total run | | 1.00 | .33 |
| Winning rate | | | 1.00 |

- We compare $\boldsymbol{C}$ and $\boldsymbol{\Sigma}$ to determine the set of unknown parameters: $\sigma_2, \sigma_3, a_{12}, a_{23}$.
- $\sigma_1$ is the standard deviation of $x_1$, and it is observable. When it is normalized: 1.00.
- We have six independent equations, hence, these parameters can be computed (solution is introduced later).

# 3.3.1 Link model

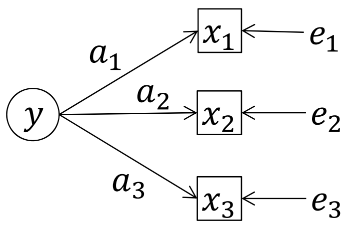



- Using the computed parameters, the 3-link model is established.
- When all the observed variables are normalized ($\sigma^2 = 1$), $R^2$ value of $x_1$ is computed by using the variance of $e_i$: $R^2 = 1 - \sigma_i^2$.

# 3.3.2 Inherent factor model

$y$: Inherent variable with mean and variance of being 0 and 1, respectively.

$$y \perp e_i$$




- Unobservable or unobserved variable is a factor to determine three observed variables.

Structural equations

$$\begin{cases} x_1 = a_1 y + e_1 \\ x_2 = a_2 y + e_2 \\ x_3 = a_3 y + e_3 \end{cases} \tag{3.9}$$

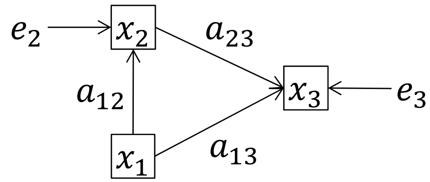
# 3.3.2 Inherent factor model

$$\boldsymbol{\Sigma} = \begin{bmatrix} a_1^2\sigma_y^2 + \sigma_1^2 & a_1 a_2 \sigma_y^2 & a_1 a_3 \sigma_y^2 \\ & a_2^2\sigma_y^2 + \sigma_2^2 & a_2 a_3 \sigma_y^2 \\ & & a_3^2\sigma_y^2 + \sigma_3^2 \end{bmatrix} \tag{3.10}$$

$$\sigma_y^2 = 1$$

Note that $y$ is an imaginary variable and its mean and variance have to be defined. Hence, unknown parameters are six: $a_1, a_2, a_3, \sigma_1, \sigma_2, \sigma_3$.

# 3.3.3 Sequential model




- $x_1$ has both direct and indirect impacts on $x_3$.

Structural equations

$$\begin{cases} x_1 = x_1 \\ x_2 = a_{12} x_1 + e_2 \\ x_3 = a_{13} x_1 + a_{23} x_2 + e_3 \end{cases} \tag{3.11}$$

# 3.3.3 Sequential model

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & a_{12}\sigma_1^2 & a_{13}\sigma_1^2 + a_{12}a_{23}\sigma_1^2 \\ & a_{12}^2\sigma_1^2 + \sigma_2^2 & a_{12}a_{13}\sigma_1^2 + a_{12}^2 a_{23}\sigma_1^2 + a_{23}\sigma_2^2 \\ & & \mathrm{Var}[x_3] \end{bmatrix}$$

$$\mathrm{Var}[x_3] = a_{13}^2\sigma_1^2 + a_{23}^2\left(a_{12}^2\sigma_1^2 + \sigma_2^2\right) + 2a_{12}a_{13}a_{23}\sigma_1^2 + \sigma_3^2$$

(3.12)

$$\sigma_1^2 = 1$$

# 3.4 General expression of covariance structure
## 3.4.1 Covariance structure of 3-link model

$$\begin{tikzcd} \boxed{x_1} \arrow[r, "a_{12}"] & \boxed{x_2} \arrow[r, "a_{23}"] & \boxed{x_3} \\ & e_2 \arrow[u] & e_3 \arrow[u] \end{tikzcd}$$

$$\begin{cases} x_1 = x_1 \\ x_2 = a_{12}x_1 + e_2 \\ x_3 = a_{23}x_2 + e_3 \end{cases} \qquad (3.1)$$

Express (3.1) in a matrix form:

Endogenous vars. (内生変数)

Exogenous vars.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ a_{12} & 0 & 0 \\ 0 & a_{13} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} x_1 \\ e_2 \\ e_3 \end{bmatrix} \qquad (3.4.1)$$

Vector of observed variables

Effect matrix

Vector of exogenous (外生) (independent) variables

# 3.4.1 Covariance structure of 3-link model

- (3.4.1) is

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e} \qquad (3.4.2)$$

- Summarizing (3.4.2) about $\boldsymbol{x}$, the vector of observed variables $\boldsymbol{x}$ is then

$$(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{e}$$

$$\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{e} \qquad (3.4.3)$$

- Use $\boldsymbol{T}$ to express $(\boldsymbol{I} - \boldsymbol{A})^{-1}$, (3.4.3) is

$$\boldsymbol{x} = \boldsymbol{T}\boldsymbol{e} \qquad (3.4.4)$$

# 3.4.1 Covariance structure of 3-link model

- The matrix of $\boldsymbol{x}$ is

$$\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] = \begin{bmatrix} \mathrm{E}[x_1^2] & \mathrm{E}[x_1x_2] & \mathrm{E}[x_1x_3] \\ & \mathrm{E}[x_2^2] & \mathrm{E}[x_2x_3] \\ & & \mathrm{E}[x_3^2] \end{bmatrix} \qquad (3.4.5)$$

- (3.4.5) is equal to $\boldsymbol{\Sigma}$ when $x_i$ variables are normalized. Because

$$\mathrm{Var}[x_i] = \mathrm{E}\left[x_i^2\right] - \mathrm{E}[x_i]^2 = \mathrm{E}\left[x_i^2\right] - 0 = \mathrm{E}\left[x_i^2\right] \qquad (3.4.6)$$

$$\mathrm{Cov}\left[x_i x_j\right] = \frac{\mathrm{E}\left[(x_i - \bar{x}_i)\left(x_j - \bar{x}_j\right)\right]}{\sqrt{\mathrm{Var}[x_i]\mathrm{Var}\left[x_j\right]}} = \mathrm{E}\left[x_i x_j\right] \qquad (3.4.7)$$

# 3.4.1 Covariance structure of 3-link model

- Hence, (3.4.8) holds:

$$\begin{aligned}\boldsymbol{\Sigma} &= \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] \\ &= \mathrm{E}[\boldsymbol{T}\boldsymbol{e}(\boldsymbol{T}\boldsymbol{e})^{\mathrm{T}}] \\ &= \mathrm{E}[\boldsymbol{T}\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{T}^{\mathrm{T}}] \\ &= \boldsymbol{T}\mathrm{E}[\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}]\boldsymbol{T}^{\mathrm{T}} \\ &= \boldsymbol{T}\boldsymbol{\Sigma}_e\boldsymbol{T}^{\mathrm{T}}\end{aligned} \tag{3.4.8}$$

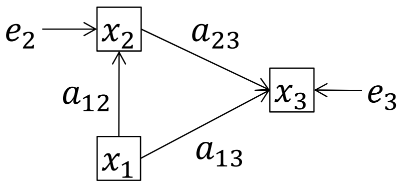
- $\boldsymbol{\Sigma}_e$ is the covariance matrix of exogenous variables.

# 3.4.1 Covariance structure of 3-link model

- $\boldsymbol{\Sigma}_e$ is

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathrm{e}} &= \mathrm{E}[\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}] \\ &= \mathrm{E}\begin{bmatrix}\mathrm{E}[x_1^2] & \mathrm{E}[x_1e_2] & \mathrm{E}[x_1e_3] \\ & \mathrm{E}[e_2^2] & \mathrm{E}[e_2e_3] \\ & & \mathrm{E}[e_3^2]\end{bmatrix} \\ &= \mathrm{E}\begin{bmatrix}\sigma_1^2 = 1 & 0 & 0 \\ & \sigma_2^2 & 0 \\ & & \sigma_3^2\end{bmatrix}\end{aligned} \tag{3.4.9}$$

# 3.4.2 Covariance structure of sequential model




Structural equations

$$\begin{cases} x_1 = x_1 \\ x_2 = a_{12}x_1 + e_2 \\ x_3 = a_{13}x_1 + a_{23}x_2 + e_3 \end{cases} \tag{3.11}$$

$$\begin{bmatrix}x_1 \\ x_2 \\ x_3\end{bmatrix} = \begin{bmatrix}0 & 0 & 0 \\ a_{12} & 0 & 0 \\ a_{13} & a_{23} & 0\end{bmatrix}\begin{bmatrix}x_1 \\ x_2 \\ x_3\end{bmatrix} + \begin{bmatrix}x_1 \\ e_2 \\ e_3\end{bmatrix} \tag{3.4.10}$$

$$\Downarrow$$

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$$

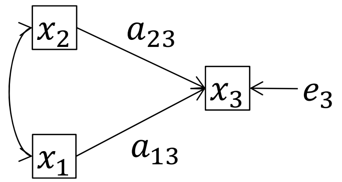
# 3.4.2 Covariance structure of sequential model

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathrm{e}} &= \mathrm{E}[\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}] \\ &= \mathrm{E}\begin{bmatrix}\sigma_1^2 = 1 & 0 & 0 \\ & \sigma_2^2 & 0 \\ & & \sigma_3^2\end{bmatrix}\end{aligned} \tag{3.4.11}$$

The covariance structure is

$$\begin{aligned}\boldsymbol{\Sigma} &= \boldsymbol{T}\boldsymbol{\Sigma}_e\boldsymbol{T}^{\mathrm{T}} \\ &= (\boldsymbol{I} - \boldsymbol{A})^{-\mathbf{1}}\boldsymbol{\Sigma}_e\left((\boldsymbol{I} - \boldsymbol{A})^{-\mathbf{1}}\right)^{\mathrm{T}}\end{aligned}$$

## 3.4.3 Covariance structure of multiple regression model




Structural equations

$$\begin{cases} x_1 = x_1 \\ x_2 = x_2 \\ x_3 = a_{13}x_1 + a_{23}x_2 + e_3 \end{cases} \tag{3.4.12}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ a_{13} & a_{23} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ e_3 \end{bmatrix} \tag{3.4.13}$$

$$\Downarrow$$

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$$

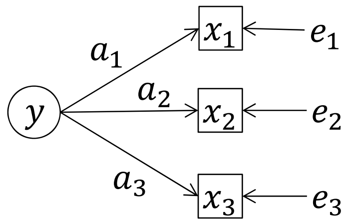
## 3.4.3 Covariance structure of multiple regression model

$$\boldsymbol{\Sigma}_{\mathrm{e}} = \mathrm{E}[\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}]$$

$$= \mathrm{E}\begin{bmatrix} \sigma_1^2 = 1 & \sigma_{12} & 0 \\ & \sigma_2^2 = 1 & 0 \\ & & \sigma_3^2 \end{bmatrix} \tag{3.4.14}$$

The covariance (correlation) between $x_1$ and $x_2$ is not zero. $\sigma_{12}$ is given from the observation.

## 3.4.4 Covariance structure of factor analysis (FA)




Structural equations

$$\begin{cases} y = y \\ x_1 = a_1 y + e_1 \\ x_2 = a_2 y + e_2 \\ x_3 = a_3 y + e_3 \end{cases} \tag{3.4.15}$$

Usually, multiple factors are considered.

Structural equations includes latent variable $y$.

$$\begin{bmatrix} y \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ a_1 & 0 & 0 & 0 \\ a_2 & 0 & 0 & 0 \\ a_3 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y \\ e_1 \\ e_2 \\ e_3 \end{bmatrix} \tag{3.4.16}$$

## 3.4.4 Covariance structure of factor analysis (FA)

$$\begin{bmatrix} y \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ a_1 & 0 & 0 & 0 \\ a_2 & 0 & 0 & 0 \\ a_3 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y \\ e_1 \\ e_2 \\ e_3 \end{bmatrix} \tag{3.4.16}$$

$$\Downarrow$$

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{bmatrix} = \boldsymbol{A}\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{bmatrix} + \boldsymbol{e} \tag{3.4.17}$$

Here,

$$\boldsymbol{y} = [y] \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

## 3.4.4 Covariance structure of factor analysis (FA)

- From (3.4.17),

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{bmatrix} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{e} = \boldsymbol{Te} \qquad (3.4.18)$$

- To convert $[\boldsymbol{y}\ \boldsymbol{x}]^{\mathrm{T}}$ to $\boldsymbol{x}$, we use $\boldsymbol{G}$ matrix:

$$\boldsymbol{x} = \boldsymbol{G}\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{bmatrix} \qquad \boldsymbol{G} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (3.4.19)$$

- The covariance structure of observed variables is

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathrm{E}[\boldsymbol{xx}^{T}] \qquad (3.4.20) \\ &= \mathrm{E}[\boldsymbol{GTe}(\boldsymbol{GTe})^{T}] \\ &= \boldsymbol{GT}\mathrm{E}[\boldsymbol{ee}^{\mathrm{T}}]\boldsymbol{T}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}} \\ &= \boldsymbol{GT\Sigma}_e\boldsymbol{T}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}} \end{aligned}$$

## 3.4.4 Covariance structure of factor analysis (FA)

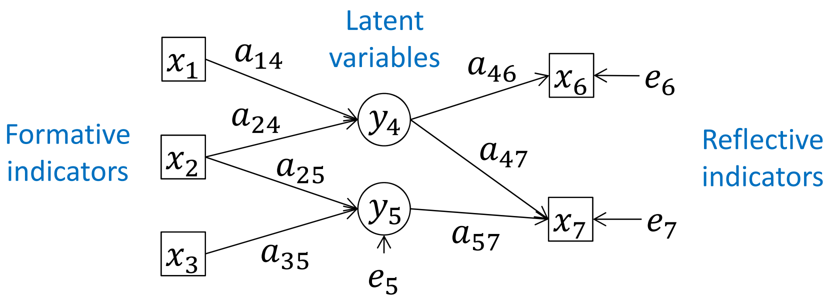
- The covariance matrix of $[\boldsymbol{y}^{\mathrm{T}}\ \boldsymbol{e}]$ is

$$\boldsymbol{\Sigma}_e = \begin{bmatrix} \sigma_y^2 = 1 & 0 & 0 & 0 \\ & \sigma_1^2 & 0 & 0 \\ & & \sigma_2^2 & 0 \\ & & & \sigma_3^2 \end{bmatrix} \qquad (3.4.21)$$

- $y$ is an imaginary variable and it does not have scales. Hence, its scale is set: $\sigma_y^2 = 1$.

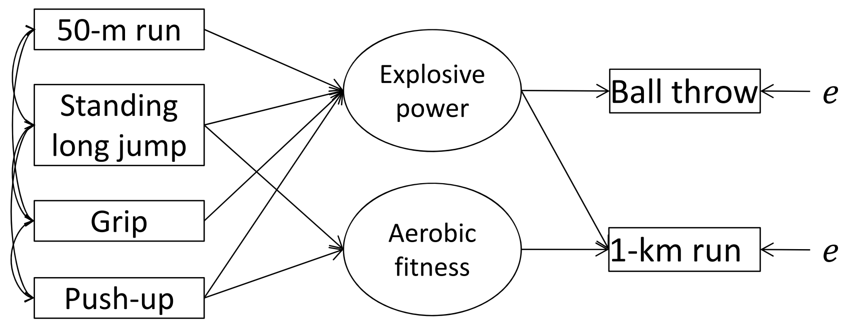
## 3.4.5 Covariance structure of MIMIC model

- Multiple indicators and multiple causes (MIMIC) model
- Effects among multiple variables (indicators) are assessed through latent variables.
- Known as a problem of formative (causal) and reflective indicators
- Mixture of PCA and FA (factor analysis)



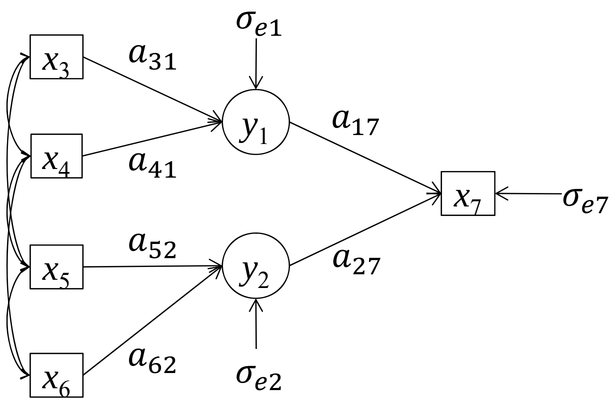

## 3.4.5 Covariance structure of MIMIC model

- Example of physical fitness test



# 3.4.5 Covariance structure of MIMIC model

- A case with four formative indicators and one reflective indicator




# 3.4.5 Covariance structure of MIMIC model

$$\begin{bmatrix} y_1 \\ y_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 0 & 0 & a_{31} & a_{41} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{52} & a_{62} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{17} & a_{27} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ e_7 \end{bmatrix} \tag{3.4.22}$$

$$\boldsymbol{G} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.4.23}$$

# 3.4.5 Covariance structure of MIMIC model

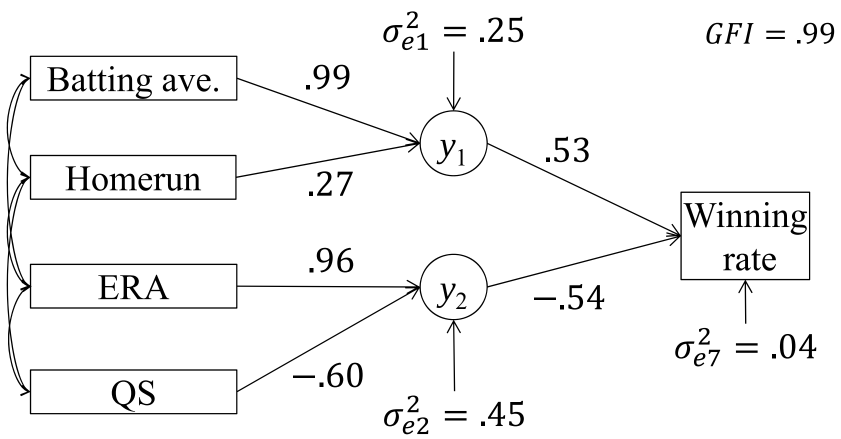
$$\boldsymbol{\Sigma}_e = \begin{bmatrix} \sigma_{e1}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ & \sigma_{e2}^2 & 0 & 0 & 0 & 0 & 0 \\ & & 1 & \sigma_{34} & \sigma_{35} & \sigma_{36} & 0 \\ & & & 1 & \sigma_{45} & \sigma_{46} & 0 \\ & & & & 1 & \sigma_{56} & 0 \\ & & & & & 1 & 0 \\ & & & & & & \sigma_{e7}^2 \end{bmatrix} \tag{3.4.24}$$

# 3.4.5 Covariance structure of MIMIC model

| | Batt ave | Homerun | ERA | QS | Win rate |
|---|---|---|---|---|---|
| Batt ave | 1 | .16 | -.00 | -.19 | .49 |
| Homerun | | 1 | .43 | -.33 | -.10 |
| ERA | | | 1 | -.67 | -.67 |
| QS | | | | 1 | .52 |
| Win rate | | | | | 1 |

$$\boldsymbol{\Sigma}_e = \begin{bmatrix} \sigma_{e1}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ & \sigma_{e2}^2 & 0 & 0 & 0 & 0 & 0 \\ & & 1 & .16 & 0 & -.19 & 0 \\ & & & 1 & .43 & -.33 & 0 \\ & & & & 1 & -.67 & 0 \\ & & & & & 1 & 0 \\ & & & & & & \sigma_{e7}^2 \end{bmatrix} \tag{3.4.25}$$

## 3.4.5 Covariance structure of MIMIC model




## 3.5 Computation/Estimation of model parameters

- The set of model parameters $\boldsymbol{\theta} = \{\boldsymbol{a},\ \boldsymbol{\sigma}\}$ is given as a generalized least squares (GLS) solution or maximum likelihood estimation (MLE).
- Here, GLS is introduced.

- The error that should be minimized is a form of matrix: $\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{S}$, here $\boldsymbol{S}$ is the observed covariance matrix.
- The cost function is

$$f(\boldsymbol{\theta}) = \mathrm{tr}[(\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{S})(\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{S})^{\mathrm{T}}] \qquad (3.4.22)$$
$$= \sum_i \sum_j (\sigma_{ij} - s_{ij})^2$$

## 3.5 Computation/Estimation of model parameters

- The cost function (3.4.22) is a polynomial equation with high-order terms. Rather than analytical approaches, we use a computer.
- Matlab/Octave provides a solver of minimization problems: fmin families.
- Let’s see an example.

## 3.5.1 link.m

```
global C;
%      x1    x2    x3
C = [  1.0   .65   .32
       .65   1.0   .49
       .32   .49   1.0];

% Optimization
x0 = rand(1,4);
[x, fval, exitflag, output] = fminunc( @mylink, x0 )
```

C is the correlation matrix of the observed variables.

X0 is the set of unknown parameters.

fmin returns the parameters which minimizes the cost function defined by mylink

## 3.5.2 mylink.m

```
function [f, x, Sg] = mylink(x)
        global C;
        Se = [  1       0       0;
                0       x(1)    0;
                0       0       x(2)    ];
        A = [   0       0       0;
                x(3)    0       0;
                0       x(4)    0       ];
        G = [   1       0       0;
                0       1       0;
                0       0       1       ];
        T = ( eye(size(A)) - A )^-1;
        Sg = G * T * Se * T' * G';
        f = trace( (C - Sg) * (C - Sg)' );
end
```

$\boldsymbol{\Sigma}_e$: Covariance matrix of exogenous variables.
$\sigma_2^2$: x(1)
$\sigma_3^2$: x(2)

$\boldsymbol{A}$: Effect matrix
$a_{12}$: x(3)
$a_{23}$: x(4)

f: Cost value

## 3.5.3 Execution of link.m

```
x =
   0.57647   0.75926   0.65039   0.49077

fval =   2.4104e-006
GFI =  0.99999

Sg =
   1.00000   0.65039   0.31920
   0.65039   0.99949   0.49052
   0.31920   0.49052   1.00000
C =
   1.00000   0.65000   0.32000
   0.65000   1.00000   0.49000
   0.32000   0.49000   1.00000
```

x: Estimated values of unknown parameters

f: Final cost value

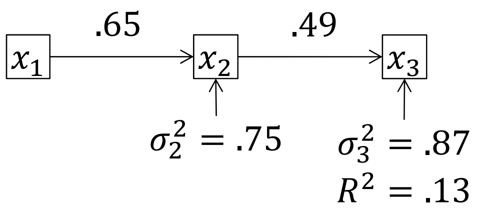
Goodness of fitting

Sg: Estimated covariance matrix

C: Observed covariance matrix

## 3.5.3 Execution of link.m

- Finally, we get the bottom link model.
- The overall fitting is good because GFI is very close to 1.0.
- However, just tiny parts of x2 and x3 are explained due to random errors.




## 3.6 Assessment of models

- As described in the introductory part, SEM does not tell us the causal relationships among the variables.
- It just tells us whether the observed variables are fit into the model statistically.
- Several indices are used. Here, GFI is introduces as one of the most representative indices.
- Other representatives include AGFI, CFI, chi-square values, etc.

## 3.6.1 Gross assessment index (GFI)

- Goodness-of-fitting (GFI) compares the similarity between the observed ($\boldsymbol{S}$) and estimated ($\boldsymbol{\Sigma}(\boldsymbol{\theta})$) covariance matrices, and is defined by

$$GFI = 1 - \frac{\mathrm{tr}[(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))^2]}{\mathrm{tr}[(\boldsymbol{S})^2]}$$

- Here, $\mathrm{tr}[\boldsymbol{A}^2] = \mathrm{tr}\left[\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right]$.
- Its maximum value is 1 when the model completely matches the data. The minimum is not zero and can be negative.
- Preferred to be greater than .90-.95.
- GFI evaluates the validity of the entire model.

## 3.6.2 Local assessment index ($R^2$)

- To see the estimation accuracy of certain variables, the coefficient of determination ($R^2$) is used.
- $R^2$ of $x_i$ is given by

$$R^2 = 1 - \frac{\sigma_{ie}^2}{\mathrm{Var}[x_i]}$$

where $\sigma_{ie}^2$ is the variance of the random error for $x_i$.

## 3.7 Direct and indirect effect

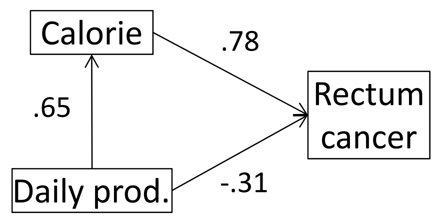
- For some cases, correlation coefficients do not indicate the direct relationship between two variables. One reason is the indirect effects.
- The following correlation matrix is calculated among the intake of calories, daily products, and the rate of rectum (直腸) cancer.
- The correlation between the daily products and rate of rectum cancer is small, which suggests a little relationship between them.

| | Daily prod. | Calories | Rectum cancer |
|---|---|---|---|
| Daily prod. | 1.0 | .65 | .20 |
| Calories | | 1.0 | .58 |
| Rectum | | | 1.0 |

[Toyoda 20111]

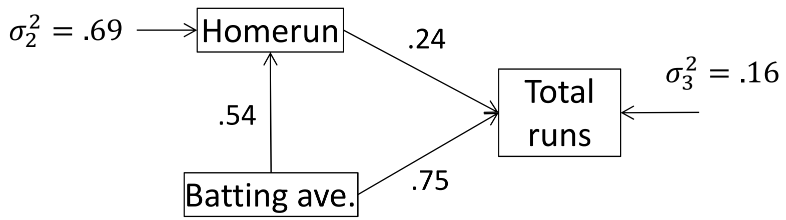
## 3.7 Direct and indirect effect

- The sequential model tells us that
  - The daily products lower the risk of rectum cancer
  - The daily products are of high calories, and then indirectly raise the risk of rectum cancer.
- If there are two men whose calorie intakes are the same, the one who eats more daily products is less at risk of rectum cancers.



# 3.7 Direct and indirect effect

- Another example from baseball.
- The relationships among the batting average, homerun, and total runs are better explained by the sequential model than a link model.
- This is fair because the batting average influences both the total runs and number of homeruns directly.




# Third report

1. Title (your name, student ID)
2. Background of data (What kind of data? Why do you want to analyze personally? Where did you get data?)
3. Choose (at least) three variables, and compare at least two models to which the data may fit.
   1. Describe which models you selected
   2. Show the observed and estimated correlation coefficients
   3. Show the estimated models with effect values, error variances, and GFI.
4. Interpret the results

Due date: July 12th



# Upcoming schedule

- June 21: 2nd class of SEM
- June 28: Covariance selection
- July 5: Your presentation
- July 12: Your presentation
- July 19: Your presentation
- July 26: Your presentation

# Final presentation

- I randomly choose ~30 students.
- Those who are selected will give their own presentations about multiple regression analysis, PCA, or SEM. You can choose.
- Each has 10 min including Q&A. The main presentation should be as long as 7-8 min.
- You may (recommended) update your presentation file for the final presentation.
- Use your own laptop or my computer for the presentation
  - Bring your USB-flash memory or you may send me your presentation files beforehand
  - Bring the HDMI or DVI adaptors for your computer