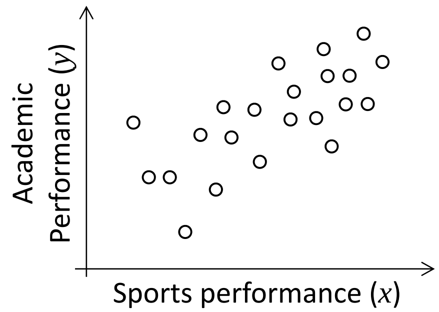
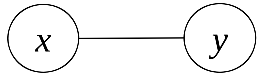
4 Covariance selection

共分散選択

4.1 Conditional independence (条件付き独立)

- In a class of an elementary school, it is observed that some pupils exhibit higher academic and physical performances.
- Can we say that academic and physical abilities are related?



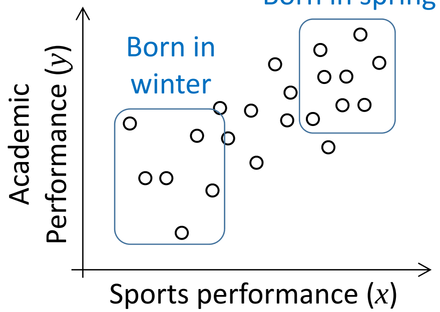


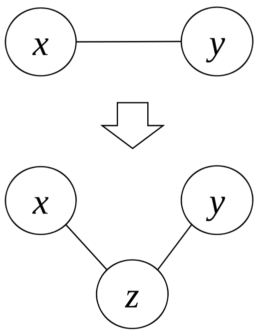
Is x related to y?

* In this chapter, we do not use bidirectional arcs.

4.1 Conditional independence (条件付き独立)

- If we look at the month of birth (z), then we realize that those born in earlier months exhibit relatively greater performances.
- x and y are directly related to z. But, they are independent for pupils born in the same season.





4.1 Conditional independence (条件付き独立)

- There is no relationship between the sports (x) and academic performances (y) provided the months of pupils (z).
- x and y are independent when the other variables are fixed. x and y are conditionally independent.
- Conditional independence is mathematically expressed as:

$$x \perp y \mid z \tag{4.1.1}$$

$$\mathrm{P}(X \cap Y|Z) = \mathrm{P}(X|Z) \times \mathrm{P}(Y|Z) \tag{4.1.2}$$

4.1 Conditional independence (条件付き独立)

- Quick review
- When two events A and B are independent, the probability of the co-occurrence is

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B) \quad (4.1.3)$$

4.2 Partial correlation coefficient (偏相関係数)

- We have p types of variables. p-2 variables are fixed and only two variables (i.e., x_i and x_j) vary.
- Under this restriction, the correlation coefficient between x_i and x_j is computed by

$$\rho_{ij.rest} = -\frac{\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}} \quad (4.1.4)$$

$$\mathbf{\Sigma} = \left(\sigma_{ij}\right), \ \ \mathbf{\Sigma}^{-1} = \left(\sigma^{ij}\right) \quad (4.1.5)$$

- $\mathbf{\Sigma}$ is the covariance matrix of p variables.
- This correlation coefficient is called the partial correlation coefficient and suggests direct relationship between two variables.

4.2 Partial correlation coefficient (偏相関係数)

$$\mathbf{\Sigma} = \left(\sigma_{ij}\right) = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ & \ddots & \vdots \\ & & \sigma_{pp} \end{bmatrix} \quad (4.1.6)$$

$$\mathbf{\Sigma}^{-1} = \left(\sigma^{ij}\right) = \begin{bmatrix} \sigma^{11} & \cdots & \sigma^{1p} \\ & \ddots & \vdots \\ & & \sigma^{pp} \end{bmatrix} \quad (4.1.7)$$

4.2 Partial correlation coefficient (偏相関係数)

- The partial correlation coefficient is computed by mathematically (virtually) fixing p-2 variables.
- Hence, it does not necessarily <u>suggest</u> the direct relationship between the two variables we concern.
- Matlab provides *partialcorr* function for the computation.

4.3 Conditional independency and partial correlation coefficient

$$\rho_{ij.rest} = 0 \quad \longleftrightarrow \quad x_i \text{ and } x_j \text{ are conditionally independent.}$$

$$\sigma^{ij} = 0$$

- When the partial correlation coefficient between a pair of variables (x_i and x_j) is zero ($\rho_{ij.rest} = 0$), these variables are conditionally independent.
- $\rho_{ij.rest} = 0$ means its numerator is zero: $\sigma^{ij} = 0$.
- Hence, the inverse covariance matrix ($\mathbf{\Sigma}^{-1}$) largely tells us the information pertaining to conditional independence.

4.4 Introduction to covariance selection

- Covariance selection is to specify the causal relationships among observed variables based on their partial correlation coefficients.
- Let's see one example to roughly understand the idea.
  - We have four variables and their correlation matrix $\mathbf{\Sigma}$ and partial correlation matrix $\boldsymbol{P}$.

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & .30 & .40 & .25 \\ & 1 & .50 & .20 \\ & & 1 & .50 \\ & & & 1 \end{bmatrix} \qquad \boldsymbol{P} = \begin{bmatrix} - & .13 & .24 & .07 \\ & - & .42 & -.08 \\ & & - & .44 \\ & & & - \end{bmatrix}$$

* We ignore the diagonal values, which are meaningless to discuss.

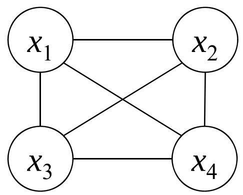
4.4 Introduction to covariance selection

- Variable pairs of very small partial correlation coefficient is nearly conditionally independent.
- Here, $\rho_{14.rest} \sim 0, \rho_{24.rest} \sim 0$.
- x_4 seems not to have direct relationships with x_1 and x_2.

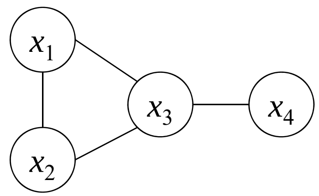
$$\boldsymbol{P} = \begin{bmatrix} - & .13 & .24 & .07 \\ & - & .42 & -.08 \\ & & - & .44 \\ & & & - \end{bmatrix}$$

4.4 Introduction to covariance selection

- The left full-model assumes the connections among all the variables: perfect graph. The full model is able to represent the observed correlation coefficients exactly.
- The right reduced model has removed supposedly independent connections. This reduced model is more concise while as accurate as the full model.



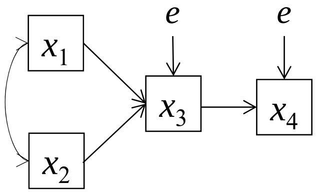
Full model.
Perfect but redundant.



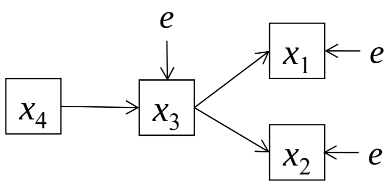
Reduced model.
Accurate and concise.

4.4 Introduction to covariance selection

- The reduced model allows us to imagine the following possible models.



Model 1



Model 2

- A reasonable model can be inferred by using the knowledge of problems. For example, when x_1, x_2, x_3, and x_4 are the scores for 100-m run, vertical jump (垂直飛び), long jump (走り幅跳び), and high jump (高跳び), respectively, then, model 1 is judged to be semantically reasonable.

4.5 Dempster's theorem

- Hat is the symbol of estimation.
- (i, j) is the connection (edge/arc) between x_i and x_j.
- $\boldsymbol{I}$ is the set of removed connections.
- $\boldsymbol{J}$ is the set of existent connections.
- $\boldsymbol{I} \cap \boldsymbol{J} = \phi$

- The maximum likelihood estimate of $\boldsymbol{\Sigma} = (\sigma_{ij})$ for which σ^{ij} $((i,j) \in \boldsymbol{I})$ are constrained to zero satisfies
  - 1) For $(i,j) \in \boldsymbol{I}$, $\hat{\sigma}^{ij} = 0$
  - 2) For $(i,j) \in \boldsymbol{J}$, $\hat{\sigma}_{ij} = \sigma_{ij}$

4.6 Procedures of covariance selection

1) Compute the correlation matrix ($\boldsymbol{\Sigma}$) from the observed data
2) Compute the inverse matrix ($\boldsymbol{\Sigma}^{-1}$)
3) Set the smallest $|\hat{\sigma}^{ij}|$ $((i,j) \in \boldsymbol{J})$ value to be zero

$$\widehat{\boldsymbol{\Sigma}}^{-1} = \begin{bmatrix} - & \sigma^{12} & \sigma^{13} & \sigma^{14} \\ & - & \sigma^{23} \to 0 & \sigma^{24} \\ & & - & \sigma^{34} \\ & & & - \end{bmatrix}$$

and add (i, j) to $\boldsymbol{I}$.

4) Compute the maximum likelihood estimate of $\widehat{\boldsymbol{\Sigma}}$ corresponding to $\widehat{\boldsymbol{\Sigma}}^{-1}$ by using Dempster's theorem
5) Compute the inverse of $\widehat{\boldsymbol{\Sigma}}$ and go to the process of 3).

4.6 Procedures of covariance selection

- Nullify the edges one by one for which partial correlation coefficient is small. This process is continued as long as the reduced (simplified) model reasonably matches $\boldsymbol{\Sigma}$, originally observed correlation matrix.
- To judge whether the reduced model statistically deviates from the full model, we use a deviance index in the next section.

4.7 Deviance index (逸脱度)

- Deviance between the full and reduced model

$$\mathrm{dev}(RM_i) = n \log \frac{|\widehat{\boldsymbol{\Sigma}}_i|}{|\boldsymbol{\Sigma}|} \qquad (4.1.8)$$

n is the number of samples used to compute $\boldsymbol{\Sigma}$.
$\widehat{\boldsymbol{\Sigma}}_i$ is the estimated correlation matrix of i-th reduced model.

- Deviance between two reduced models

$$\mathrm{dev}(RM_i, RM_{i+1}) = \mathrm{dev}(RM_{\mathrm{i+1}}) - \mathrm{dev}(RM_{\mathrm{i}}) \qquad (4.1.9)$$
$$= n \log \frac{|\widehat{\boldsymbol{\Sigma}}_{i+1}|}{|\widehat{\boldsymbol{\Sigma}}_i|}$$

4.7 Deviance index (逸脱度)

- The deviance index asymptotically follows χ^2 distribution of which d.o.f. is
  - For $\mathrm{dev}(RM_i)$, d.o.f. is the number of removed edges.
  - For $\mathrm{dev}(RM_i, RM_{i+1})$, d.o.f. is 1.

- The process of 4.7 is continued as long as the two types of χ^2 values are small.
- There is no standard for χ^2 values, however, the p value of greater than 0.5 is accepted typically. If p value is smaller than 0.5, then the process to reduce the model is stopped.

4.8 Example of track and filed

- Relationships among track-and-field sports scores: times for the 100-m-run, 400-m-run, and 1500-m-run, and the distance of long-jump（走り幅跳び）, $n = 50$. [Miyakawa 2013]
- The correlation matrix and its inverse are as follows.

Correlation matrix ($\boldsymbol{\Sigma}_0$)

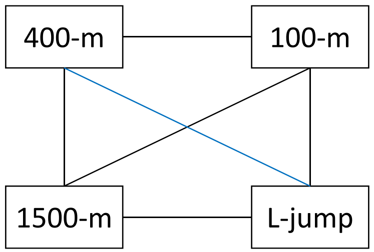
	100-m	400-m	1500-m	L-jump
100-m	1	.47	-.18	-.22
400-m		1	.21	-.12
1500-m			1	-.05
L-jump				1

$\boldsymbol{\Sigma}_0^{-1}$

	100-m	400-m	1500-m	L-jump
100-m	1.50	-.77	.44	.26
400-m		1.45	-.44	-.016
1500-m			1.18	.10
L-jump				1.06

4.8 Example of track and filed

- We start from the full model: FM.



Full model: FM

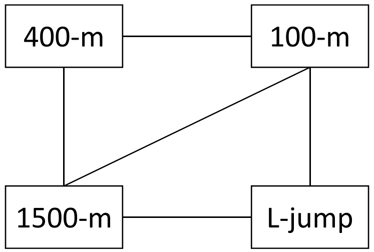
$\boldsymbol{\Sigma}_0^{-1}$

	100-m	400-m	1500-m	L-jump
100-m	1.50	-.77	.44	.26
400-m		1.45	-.44	-.016
1500-m			1.18	.10
L-jump				1.06

4.8 Example of track and filed

- We start from the full model: FM.
- Set the smallest $|\sigma^{ij}|$ value to be zero: $\hat{\sigma}_{(1)}^{24} = 0, \boldsymbol{I}_1 = \{(2,4)\}$.
- We call the new model as RM_1.

$\hat{\sigma}_{(r)}$: Value for r-th reduced model.



Reduced model: RM_1

$\boldsymbol{\Sigma}_1^{-1}$

	100-m	400-m	1500-m	L-jump
100-m	1.50	-.77	.44	.26
400-m		1.45	-.44	.00
1500-m			1.18	.10
L-jump				1.06

4.8 Example of track and filed

- We then compute $\boldsymbol{\Sigma}_1$ which is the maximum likelihood estimate of the correlation matrix corresponding to $\boldsymbol{\Sigma}_1^{-1}$.
- Only $\hat{\sigma}_{ij}$ ($(i,j) \in \boldsymbol{I}$) is modified. $\boldsymbol{I}_1 = \{(2,4)\}$.

Correlation matrix ($\boldsymbol{\Sigma}_1$)

	100-m	400-m	1500-m	L-jump
100-m	1	.47	-.18	-.22
400-m		1	.21	-.13
1500-m			1	-.05
L-jump				1

$$\hat{\sigma}_{(1)24} = \sigma_{(0)24} + \frac{\sigma_{(0)}^{24}}{\sigma_{(0)}^{22}\sigma_{(0)}^{44} - (\sigma_{(0)}^{24})^2}$$

$$-.13 = -.12 + \frac{-.016}{1.45 \times 1.06 - (-.016)^2} \tag{4.8.1}$$

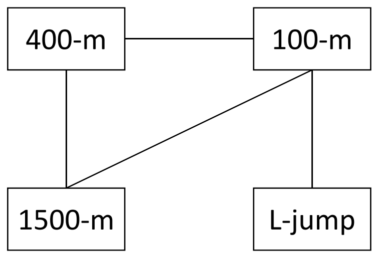
$\hat{\sigma}_{(r)}$: Value for r-th reduced model.

4.8 Example of track and filed

- $\mathrm{dev}(RM_1) = 0.0088, p = 0.93$.
  - FM and RM_1 are statistically similar.
- We then accept RM_1 and move to the 2nd reduced model.
  - $\boldsymbol{I}_1 = \{(2,4)\}$
  - $\boldsymbol{J}_1 = \{(1,2), (1,3), (1,4), (2,3), (3,4)\}$

4.8 Example of track and filed

- Set the smallest edge ($\in \boldsymbol{J}_1$) in $\boldsymbol{\Sigma}_1^{-1}$ to be zero.
- New $\boldsymbol{\Sigma}^{-1}$ is called $\boldsymbol{\Sigma}_2^{-1}$.
- For RM_2 , $\boldsymbol{I}_2 = \{(2,4), (3,4)\}$.



Reduced model: RM_2

$\boldsymbol{\Sigma}_2^{-1}$

	100-m	400-m	1500-m	L-jump
100-m	1.49	-.76	.44	.25
400-m		1.45	-.44	.00
1500-m			1.18	.10→ .00
L-jump				1.06

4.8 Example of track and filed

- $\mathbf{\Sigma}_2$ is computed from $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma}_1^{-1}$.
- $\hat{\sigma}_{(2)24}$ and $\hat{\sigma}_{(2)34}$ are updated because $\boldsymbol{I}_2 = \{(2,4),(3,4)\}$.

$$(1) \qquad \hat{\sigma}_{(2)24} = \hat{\sigma}_{(1)24} + \frac{\hat{\sigma}_{(1)}^{24}}{\hat{\sigma}_{(1)}^{22}\hat{\sigma}_{(1)}^{44} - (\hat{\sigma}_{(1)}^{24})^2}$$

$$(2) \qquad \hat{\sigma}_{(2)34} = \hat{\sigma}_{(1)34} + \frac{\hat{\sigma}_{(1)}^{34}}{\hat{\sigma}_{(1)}^{33}\hat{\sigma}_{(1)}^{44} - (\hat{\sigma}_{(1)}^{34})^2}$$

$$(3) \qquad \Sigma_1^{-1} \leftarrow \Sigma_2^{-1}$$

- (1)—(3) are repeated until $\hat{\sigma}_{(2)24}$ and $\hat{\sigma}_{(2)34}$ converge.
- Algorithm of N. Wermuth and E. Scheidt (1977).

4.8 Example of track and filed

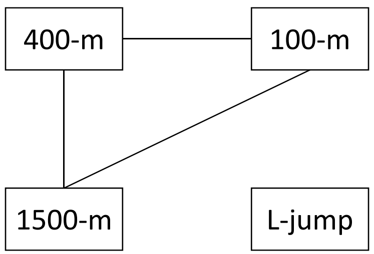
- $\mathbf{\Sigma}_2$ is determined as follows.
  - $\mathrm{dev}(RM_2) = 0.45, p = 0.80 > 0.50.$
  - $\mathrm{dev}(RM_2, RM_1) = 0.43, p = 0.51 > 0.50.$
- We then accept RM_2 and move to the 2nd reduced model.
  - $\boldsymbol{I}_2 = \{(2,4),(3,4)\}$
  - $\boldsymbol{J}_2 = \{(1,2),(1,3),(1,4),(2,3)\}$

Correlation matrix ($\mathbf{\Sigma}_2$)

	100-m	400-m	1500-m	L-jump
100-m	1	.47	-.18	-.22
400-m		1	.21	-.10
1500-m			1	.04
L-jump				1

4.8 Example of track and filed

- Set the smallest edge ($\in \boldsymbol{J}_2$) in $\mathbf{\Sigma}_2^{-1}$ to be zero.
- New $\mathbf{\Sigma}^{-1}$ is called $\mathbf{\Sigma}_3^{-1}$.
- For RM_2 , $\boldsymbol{I}_2 = \{(1,4),(2,4),(3,4)\}$.



Reduced model: RM_3

$\mathbf{\Sigma}_3^{-1}$

	100-m	400-m	1500-m	L-jump
100-m	1.49	-.76	.42	.23→ .00
400-m		1.45	-.44	.00
1500-m			1.17	.00
L-jump				1.05

4.8 Example of track and filed

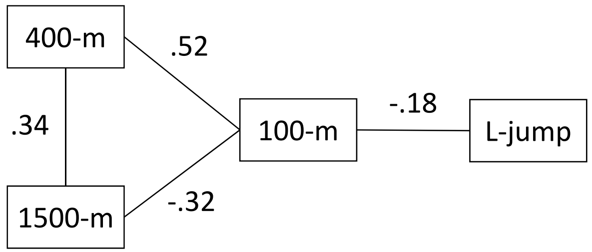
- Using the algorithm of Wermuth and E. Scheidt, $\mathbf{\Sigma}_3$ is determined as follows.
  - $\mathrm{dev}(RM_3) = 2.92, p = 0.40 < 0.50.$
  - $\mathrm{dev}(RM_3, RM_2) = 2.48, p = 0.12 < 0.50.$
- We then reject RM_3.

Correlation matrix ($\mathbf{\Sigma}_3$)

	100-m	400-m	1500-m	L-jump
100-m	1	.47	-.18	.00
400-m		1	.21	.00
1500-m			1	.00
L-jump				1

4.8 Example of track and filed

- We finally accept RM_2.
- The-100-m and 400-m runs positively influence each other while the-100-m and 1500-m runs negatively do so.
- Those with the good records of the long-jump are good sprinters. Good jumpers are good sprinters.



* Values are the partial correlation coefficients.

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. Draw the estimated models with effect values, error variances, and GFI.
4. Interpret the results

Due date: July 12th

Presentation schedule

- July 5
    - Mr. Hasegawa, Mr. Schatz, Mr. Y. Song, Mr. K. Inoue, Mr. K. Iwata, Mr. Kito, Ms. 許, and Ms. Wu
- July 12
    - Mr. Ujiie, Mr. Hara, Mr. Hongyuan, Mr. H. Iwai, Mr. Ura, Mr. Hongyuan, and Mr. Qiu
- July 19
    - Ms. Luthardt, Mr. Kanada, Mr. R. Inoue, Mr. Naito, Ms. F. Li, Mr. T. Sato, and Mr. Miyahara.
- July 26
    - Ms. Ji, Mr. Sennin, Mr. Taga, Mr. Imagawa, + a few more

Final presentation

- I randomly chose ~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