

**พีชคณิตเมทริกซ์
และหลักความเป็นไปได้สูงสุด
(Matrix Algebra and
Maximum Likelihood)**



สันศักดิ์ พรประเสริฐมานิต



โครงร่างนำเสนอ

- ลักษณะของเมทริกซ์
- การจัดการเมทริกซ์
- เมทริกซ์ในการวิเคราะห์ข้อมูล
- ค่าคาดหวังและความแปรปรวนร่วม
- พังค์ชันเพิ่มเติม
- หลักการความเป็นไปได้สูงสุด



ลักษณะของเมทริกซ์

- เมทริกซ์เป็นการจัดตัวเลขในรูปของสี่เหลี่ยม ที่มีแถวและคอลัมน์

คอลัมน์มีทั้งหมด n คอลัมน์

แถวมีทั้งหมด m แถว

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = (a_{ij})_{m \times n}$$

- เมทริกซ์ \mathbf{A} มีขนาด m แถว n คอลัมน์ หรือเรียกว่าเมทริกซ์ \mathbf{A} มีขนาด $m \times n$
- a_{ij} หมายถึงสมาชิกของเมทริกซ์แถวที่ i คอลัมน์ที่ j

ลักษณะของเมทริกซ์

- กรณีที่เมทริกซ์มีคอลัมน์เดียว อาจเรียกว่าเวกเตอร์ (Vector) เรียกเต็มๆ ว่าเวกเตอร์คอลัมน์ (Column vector) จะใช้ตัวอักษรภาษาอังกฤษเล็กหนา แทนสัญลักษณ์

$$\mathbf{b} = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{m1} \end{bmatrix} = (b_{ij})_{m \times 1}$$

แถวมีทั้งหมด m แถว

- เวกเตอร์ \mathbf{b} มีสมาชิก m ตัว

การจัดการเมทริกซ์



- เมทริกซ์หรือเวกเตอร์จะมีค่าเท่ากัน เมื่อมีขนาดเท่ากัน และสมาชิกทุกตัวมีค่าเท่ากัน
- ให้ $\mathbf{A} = (a_{ij})_{m \times n}$ และ $\mathbf{B} = (b_{ij})_{m \times n}$ แล้ว $\mathbf{A} = \mathbf{B}$ ก็ต่อเมื่อ $a_{ij} = b_{ij}$ ในทุกค่าของ i และ j



การจัดการเมทริกซ์

- ผลรวมหรือผลต่างระหว่างเมทริกซ์หรือเวกเตอร์ จะเท่ากับผลรวมหรือผลต่างของสมาชิกแต่ละตัว
- หาผลรวมหรือผลต่างได้ ก็ต่อเมื่อเมทริกซ์หรือเวกเตอร์ทั้งสองตัวมีขนาดเท่ากัน
- ให้ $\mathbf{A} = (a_{ij})_{m \times n}$ และ $\mathbf{B} = (b_{ij})_{m \times n}$
- $\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij})_{m \times n}$ และ $\mathbf{A} - \mathbf{B} = (a_{ij} - b_{ij})_{m \times n}$
- เช่น

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ 4 & -2 \\ 6 & -8 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 7 & 2 \\ 11 & -2 \end{bmatrix}$$

การจัดการเมทริกซ์

- เมทริกซ์ศูนย์ (Zero matrix) คือ เมทริกซ์ที่สมาชิกทุกตัวเป็น 0

$$\mathbf{0}_{2 \times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \mathbf{0}_{3 \times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

- เมทริกซ์ศูนย์เป็นเมทริกซ์ที่ไปบวกหรือลบเมทริกซ์ใด จะทำให้ได้เมทริกซ์เดิม

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

การจัดการเมทริกซ์

- การคูณด้วยค่าคงที่ (Scalar Multiplication) คือ การนำค่าคงที่ไปคูณสมาชิกทุกตัวในเมทริกซ์
- $c\mathbf{A}_{m \times n} = (c \cdot a_{ij})_{m \times n}$
- เช่น

$$2 \cdot \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \\ 10 & 12 \end{bmatrix}$$

การจัดการเมทริกซ์

- ทรานโพส (Transpose) คือ การกลับเมทริกซ์จากคอลัมน์เป็นแถว และแถวเป็นคอลัมน์ ใช้เครื่องหมาย ' หรือ T อยู่ขวาบนของเมทริกซ์
- ให้ $\mathbf{A} = (a_{ij})_{m \times n}$ แล้ว $\mathbf{A}' = (a_{ji})_{n \times m}$

- เช่น

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \mathbf{A}' = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

- \mathbf{A} จะมีขนาด 3×2 แต่ \mathbf{A}' จะมีขนาด 2×3

การจัดการเมทริกซ์

- การคูณเมทริกซ์ (Matrix Multiplication) จะค่อนข้างซับซ้อน
- ให้ **AB** เป็นการคูณระหว่างเมทริกซ์ **A** และเมทริกซ์ **B** โดย **A** จะอยู่ในอันดับก่อนหน้าเมทริกซ์ **B** (ลำดับมีความสำคัญในการคูณเมทริกซ์ ไม่สามารถสลับได้)
- เมทริกซ์จะคูณกันได้ เมื่อจำนวนคอลัมน์ของเมทริกซ์แรก เท่ากับจำนวนแถวของเมทริกซ์ที่สอง

$$\mathbf{A}_{m \times n} \quad \mathbf{B}_{n \times p}$$

มีตัวเลขเหมือนกัน

- ขนาดของผลลัพธ์ จะมีจำนวนแถวเท่ากับเมทริกซ์แรก และจำนวนคอลัมน์เท่ากับเมทริกซ์หลัง ให้ **C = AB** แล้ว **C** จะมีขนาดเท่ากับ $m \times p$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} & a_{11}b_{14} + a_{12}b_{24} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} & a_{21}b_{14} + a_{22}b_{24} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} & a_{31}b_{14} + a_{32}b_{24} \end{bmatrix}$$

สมาชิกแต่ละตัวของ \mathbf{AB} คือ ผลรวมเชิงเส้นของแต่ละแถวใน \mathbf{A} และแต่ละคอลัมน์ใน \mathbf{B}

ให้ $\mathbf{C} = \mathbf{AB}$ แล้ว $c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0 & 2 & -1 \\ -1 & 0 & 2 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 1 \cdot 0 + 2 \cdot -1 & 1 \cdot 2 + 2 \cdot 0 & 1 \cdot -1 + 2 \cdot 2 \\ 3 \cdot 0 + 4 \cdot -1 & 3 \cdot 2 + 4 \cdot 0 & 3 \cdot -1 + 4 \cdot 2 \end{bmatrix} = \begin{bmatrix} -2 & 2 & 3 \\ -4 & 6 & 5 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 2 \\ 3 & -1 & 4 \\ -2 & -3 & -4 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 0.5 \\ -0.5 \end{bmatrix}$$

$$\mathbf{Py} = \begin{bmatrix} 1 \cdot 1 + 0 \cdot 0.5 + 2 \cdot -0.5 \\ 3 \cdot 1 + (-1) \cdot 0.5 + 4 \cdot -0.5 \\ (-2) \cdot 1 + (-3) \cdot 0.5 + (-4) \cdot -0.5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.5 \\ -1.5 \end{bmatrix}$$

การจัดการเมทริกซ์

- เมทริกซ์รูปสี่เหลี่ยมจัตุรัส (Square Matrix) คือเมทริกซ์ที่มีจำนวนแถวเท่ากับจำนวนคอลัมน์ เช่น

$$\mathbf{A}_{2 \times 2} = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} \quad \mathbf{B}_{3 \times 3} = \begin{bmatrix} 5 & 2 & 3 \\ 7 & 12 & 5 \\ -7 & -2 & 11 \end{bmatrix}$$

- ในเมทริกซ์แบบสมมาตร สมาชิกบนแนวทแยง (On-diagonal elements) คือ สมาชิกที่อยู่บนแนวซ้ายบนไปขวาล่าง ($i = j$) ส่วนสมาชิกนอกแนวทแยง (Off-diagonal elements) คือ สมาชิกที่ไม่ได้อยู่บนเส้นซ้ายบนไปขวาล่าง ($i \neq j$)

การจัดการเมทริกซ์

- เมทริกซ์สมมาตร (Symmetric Matrix) คือ เมทริกซ์ที่สมาชิกแถวที่ i คอลัมน์ที่ j มีค่าเท่ากับสมาชิกแถวที่ j คอลัมน์ที่ i ทุกตัว หรือ $a_{ij} = a_{ji}$
- เมทริกซ์สมมาตรต้องเป็นเมทริกซ์สี่เหลี่ยมจัตุรัสเท่านั้น

$$\mathbf{A}_{1 \times 1} = [1]$$

$$\mathbf{B}_{2 \times 2} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

$$\mathbf{C}_{3 \times 3} = \begin{bmatrix} 5 & 2 & -7 \\ 2 & 12 & 5 \\ -7 & 5 & 11 \end{bmatrix}$$

$$\mathbf{D}_{4 \times 4} = \begin{bmatrix} 5 & 2 & -7 & -1 \\ 2 & 12 & 5 & 8 \\ -7 & 5 & 11 & 9 \\ -1 & 8 & 9 & 0 \end{bmatrix}$$

$$\mathbf{E}_{5 \times 5} = \begin{bmatrix} 1 & -1 & -2 & -3 & -4 \\ -1 & 2 & -5 & -6 & -7 \\ -2 & -5 & 3 & -8 & -9 \\ -3 & -6 & -8 & 4 & -10 \\ -4 & -7 & -9 & -10 & 5 \end{bmatrix}$$

การจัดการเมทริกซ์

- เมทริกซ์แนวทแยงมุม (Diagonal Matrix) คือเมทริกซ์ที่สมาชิกนอกแนวทแยงมีค่าเท่ากับ 0

$$\mathbf{D}_{1 \times 1} = [8]$$

$$\mathbf{D}_{2 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

$$\mathbf{D}_{3 \times 3} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -12 & 0 \\ 0 & 0 & 11 \end{bmatrix}$$

การจัดการเมทริกซ์

- เมทริกซ์เอกลักษณ์ (Identity Matrix) คือเมทริกซ์ที่สมาชิกบนแนวทแยงเท่ากับ 1 และนอกแนวทแยงมีค่าเท่ากับ 0

$$\mathbf{I}_{1 \times 1} = [1]$$

$$\mathbf{I}_{2 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{I}_{3 \times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{I}_{4 \times 4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{I}_{5 \times 5} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

การจัดการเมทริกซ์

- เมทริกซ์เอกลักษณ์ จะสามารถคูณกับเมทริกซ์ใด ก็ได้เมทริกซ์เดิม ทั้งคูณข้างหน้า และคูณด้านหลัง

$$\mathbf{A}_{3 \times 2} \mathbf{I}_{2 \times 2} = \mathbf{I}_{3 \times 3} \mathbf{A}_{3 \times 2} = \mathbf{A}_{3 \times 2}$$

- อินเวอร์สของเมทริกซ์ (Matrix Inverse) คือ เมทริกซ์ใดก็ตาม ที่มาคูณแล้วได้เมทริกซ์เอกลักษณ์ ในที่นี้จะจำกัดเฉพาะเมทริกซ์สี่เหลี่ยมจัตุรัส

$$\mathbf{A}_{n \times n} \mathbf{A}_{n \times n}^{-1} = \mathbf{A}_{n \times n}^{-1} \mathbf{A}_{n \times n} = \mathbf{I}_{n \times n}$$

การจัดการเมทริกซ์

- เช่น

```
> A <- matrix(c(9, 2, 3,  
+             4, 5, 6,  
+             7, 8, 9), 3, 3, byrow=TRUE)
```

```
> Ainv <- solve(A)
```

```
> A
```

```
      [,1] [,2] [,3]  
[1,]    9    2    3  
[2,]    4    5    6  
[3,]    7    8    9
```

`solve` คือคำสั่งในการหา Inverse ของเมทริกซ์

```
> Ainv
```

```
      [,1]      [,2]      [,3]  
[1,]  0.125 -0.250000  0.125000  
[2,] -0.250 -2.500000  1.750000  
[3,]  0.125  2.416667 -1.541667
```

`%*%` เป็นคำสั่งทำ matrix multiplication

```
> round(A %*% Ainv, 3)
```

```
      [,1] [,2] [,3]  
[1,]    1    0    0  
[2,]    0    1    0  
[3,]    0    0    1
```

$$AA^{-1} = I$$

```
> round(Ainv %*% A, 3)
```

```
      [,1] [,2] [,3]  
[1,]    1    0    0  
[2,]    0    1    0  
[3,]    0    0    1
```

$$A^{-1}A = I$$

เมทริกซ์ในการวิเคราะห์ข้อมูล

- ข้อมูลแท้จริงแล้ว ก็อยู่ในรูปของเมทริกซ์ สมมติว่ามีกลุ่มตัวอย่าง N คน มีตัวแปร p ตัวแปร

$$\mathbf{X}_{N \times p} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1p} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2p} \\ X_{31} & X_{32} & X_{33} & \cdots & X_{3p} \\ X_{41} & X_{42} & X_{43} & \cdots & X_{4p} \\ X_{51} & X_{52} & X_{53} & \cdots & X_{5p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & X_{N3} & \cdots & X_{Np} \end{bmatrix}$$

```
> as.matrix(dat)
      x1 x2 x3 x4 x5
[1,]  2  2  2  3  3
[2,]  3  3  2  3  2
[3,]  3  3  3  2  3
[4,]  2  3  2  2  2
[5,]  2  2  2  3  3
[6,]  1  2  1  2  3
[7,]  2  2  2  2  1
[8,]  3  2  2  3  2
[9,]  2  2  2  3  2
[10,] 2  3  3  3  2
[11,] 1  2  2  3  2
[12,] 3  2  2  2  1
[13,] 3  2  1  3  2
[14,] 1  2  1  2  2
[15,] 3  4  3  2  2
[16,] 2  3  3  2  2
[17,] 3  2  3  2  3
[18,] 2  4  3  3  2
[19,] 2  3  3  2  3
[20,] 3  2  3  3  2
[21,] 1  3  1  2  2
[22,] 2  3  2  1  2
[23,] 2  3  2  2  1
[24,] 3  2  1  2  2
[25,] 2  3  2  2  1
[26,] 3  3  3  3  3
[27,] 2  2  2  2  3
[28,] 2  3  2  2  3
[29,] 2  2  2  2  2
[30,] 2  3  4  3  2
```

เมทริกซ์ในการวิเคราะห์ข้อมูล

- เวกเตอร์ค่าเฉลี่ย (Mean vector) คือ เวกเตอร์ที่มีค่าเฉลี่ยของแต่ละตัวแปร

$$\mathbf{m}_{p \times 1} = \frac{1}{N} (\mathbf{X}'_{N \times p} \mathbf{1}_{N \times 1})$$

โดย $\mathbf{1}$ คือ เวกเตอร์ที่มีสมาชิกทั้งหมดเป็น 1

```
> X <- as.matrix(dat)
> one_n <- as.matrix(rep(1, nrow(X)))
> m <- (t(X) %*% one_n) / nrow(X)
> m
      [,1]
x1 2.225
x2 2.470
x3 2.050
x4 2.225
x5 2.110
> as.matrix(colMeans(X))
      [,1]
x1 2.225
x2 2.470
x3 2.050
x4 2.225
x5 2.110
```

เมทริกซ์ในการวิเคราะห์ข้อมูล

- เมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) คือ เมทริกซ์ที่สมาชิกแนวทแยงเป็นความแปรปรวน (Variance) ของแต่ละตัวแปร และสมาชิกนอกแนวทแยงเป็นความแปรปรวนร่วม (Covariance) ของตัวแปรแต่ละคู่

$$\mathbf{S} = \begin{bmatrix} \text{Var}(x_1) & & & & \\ \text{Cov}(x_2, x_1) & \text{Var}(x_2) & & & \\ \text{Cov}(x_3, x_1) & \text{Cov}(x_3, x_2) & \text{Var}(x_3) & & \\ \vdots & \vdots & \vdots & \ddots & \\ \text{Cov}(x_p, x_1) & \text{Cov}(x_p, x_2) & \text{Cov}(x_p, x_3) & \cdots & \text{Var}(x_p) \end{bmatrix} = \begin{bmatrix} s_1^2 & & & & \\ s_{21} & s_2^2 & & & \\ s_{31} & s_{32} & s_3^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ s_{p1} & s_{p2} & s_{p3} & \cdots & s_p^2 \end{bmatrix}$$

เมทริกซ์ในการวิเคราะห์ข้อมูล

- เมทริกซ์ความแปรปรวนร่วม สามารถคำนวณได้จากพีชคณิตเมทริกซ์เช่นกัน

$$S = \frac{1}{N-1} \left[(\mathbf{X} - \mathbf{1}_{n \times 1} \mathbf{m}'_{p \times 1})' (\mathbf{X} - \mathbf{1}_{n \times 1} \mathbf{m}'_{p \times 1}) \right]$$

โดย $\mathbf{X} - \mathbf{1}_{n \times 1} \mathbf{m}'_{p \times 1}$ คือ คะแนนดิบที่ถูกย้ายศูนย์กลางมาอยู่ที่ค่าเฉลี่ย (Deviation Score)

```
> one_n <- matrix(1, nrow(X),1)
> dev <- X - one_n %*% t(m)
> (t(dev) %*% dev)/(nrow(X)-1)
      x1      x2      x3      x4      x5
x1 0.47675879 0.09974874 0.1896985 0.04962312 0.09572864
x2 0.09974874 0.47145729 0.2829146 0.20025126 0.07869347
x3 0.18969849 0.28291457 0.6105528 0.18467337 0.10000000
x4 0.04962312 0.20025126 0.1846734 0.45665829 0.14095477
x5 0.09572864 0.07869347 0.1000000 0.14095477 0.43005025
> cov(X)
      x1      x2      x3      x4      x5
x1 0.47675879 0.09974874 0.1896985 0.04962312 0.09572864
x2 0.09974874 0.47145729 0.2829146 0.20025126 0.07869347
x3 0.18969849 0.28291457 0.6105528 0.18467337 0.10000000
x4 0.04962312 0.20025126 0.1846734 0.45665829 0.14095477
x5 0.09572864 0.07869347 0.1000000 0.14095477 0.43005025
```

เมทริกซ์ในการวิเคราะห์ข้อมูล

- เมทริกซ์สหสัมพันธ์ (Correlation Matrix) คือ การนำความแปรปรวนร่วมไปหารด้วยส่วนเบี่ยงเบนมาตรฐานของตัวแปรที่เกี่ยวข้องทั้งสองตัวแปร

$$\mathbf{R} = \begin{bmatrix} 1 & & & & & \\ \text{Cor}(x_2, x_1) & 1 & & & & \\ \text{Cor}(x_3, x_1) & \text{Cor}(x_3, x_2) & 1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ \text{Cor}(x_p, x_1) & \text{Cor}(x_p, x_2) & \text{Cor}(x_p, x_3) & \cdots & 1 & \end{bmatrix} = \begin{bmatrix} 1 & & & & & \\ r_{21} & 1 & & & & \\ r_{31} & r_{32} & 1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 & \end{bmatrix}$$

เมทริกซ์ในการวิเคราะห์ข้อมูล

- เมทริกซ์สหสัมพันธ์ (Correlation Matrix) สามารถคำนวณจากเมทริกซ์ความแปรปรวนร่วมได้เช่นกัน ก่อนอื่นต้องเข้าใจฟังก์ชันที่สกัดเฉพาะสมาชิกแนวทแยงมุม

$$\text{Diag}(\mathbf{S}) = \mathbf{D}_S = \begin{bmatrix} s_1^2 & & & & \\ 0 & s_2^2 & & & \\ 0 & 0 & s_3^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & s_p^2 \end{bmatrix} \quad \mathbf{D}_S^{-1} = \begin{bmatrix} 1/s_1^2 & & & & \\ 0 & 1/s_2^2 & & & \\ 0 & 0 & 1/s_3^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & 1/s_p^2 \end{bmatrix}$$

$$\mathbf{D}_S^{-1/2} = \begin{bmatrix} \sqrt{1/s_1^2} & & & & \\ 0 & \sqrt{1/s_2^2} & & & \\ 0 & 0 & \sqrt{1/s_3^2} & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & \sqrt{1/s_p^2} \end{bmatrix} = \begin{bmatrix} 1/s_1 & & & & \\ 0 & 1/s_2 & & & \\ 0 & 0 & 1/s_3 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & 1/s_p \end{bmatrix}$$

เมทริกซ์ในการวิเคราะห์ข้อมูล

- นำเมทริกซ์ที่เป็นการหารส่วนเบี่ยงเบนมาตรฐาน มาคูณหน้าหลังกับเมทริกซ์ความแปรปรวนร่วม เพื่อให้ได้เมทริกซ์สหสัมพันธ์

$$\mathbf{R} = \mathbf{D}_S^{-1/2} \mathbf{S} \mathbf{D}_S^{-1/2} = \begin{bmatrix} 1/s_1 & & & & \\ 0 & 1/s_2 & & & \\ 0 & 0 & 1/s_3 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & 1/s_p \end{bmatrix} \begin{bmatrix} s_1^2 & & & & \\ s_{21} & s_2^2 & & & \\ s_{31} & s_{32} & s_3^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ s_{p1} & s_{p2} & s_{p3} & \cdots & s_p^2 \end{bmatrix} \begin{bmatrix} 1/s_1 & & & & \\ 0 & 1/s_2 & & & \\ 0 & 0 & 1/s_3 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & 1/s_p \end{bmatrix}$$

$$\mathbf{R} = \mathbf{D}_S^{-1/2} \mathbf{S} \mathbf{D}_S^{-1/2} = \begin{bmatrix} s_1^2/s_1s_1 & & & & \\ s_{21}/s_2s_1 & s_2^2/s_2s_2 & & & \\ s_{31}/s_3s_1 & s_{32}/s_3s_2 & s_3^2/s_3s_3 & & \\ \vdots & \vdots & \vdots & \ddots & \\ s_{p1}/s_ps_1 & s_{p2}/s_ps_2 & s_{p3}/s_ps_3 & \cdots & s_p^2/s_ps_p \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ r_{21} & 1 & & & \\ r_{31} & r_{32} & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 \end{bmatrix}$$

```
> S <- cov(X)
> D <- diag(diag(S))
> D2 <- solve(D)^(1/2)
> D2 %*% S %*% D2
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.0000000 0.2103957 0.3516033 0.1063504 0.2114136
[2,] 0.2103957 1.0000000 0.5273179 0.4315771 0.1747664
[3,] 0.3516033 0.5273179 1.0000000 0.3497414 0.1951546
[4,] 0.1063504 0.4315771 0.3497414 1.0000000 0.3180713
[5,] 0.2114136 0.1747664 0.1951546 0.3180713 1.0000000
> cor(X)
      x1      x2      x3      x4      x5
x1 1.0000000 0.2103957 0.3516033 0.1063504 0.2114136
x2 0.2103957 1.0000000 0.5273179 0.4315771 0.1747664
x3 0.3516033 0.5273179 1.0000000 0.3497414 0.1951546
x4 0.1063504 0.4315771 0.3497414 1.0000000 0.3180713
x5 0.2114136 0.1747664 0.1951546 0.3180713 1.0000000
```

เมทริกซ์ในการวิเคราะห์ข้อมูล

- การวิเคราะห์ถดถอยแบบพหุ ก็สามารถเขียนในรูปของสมการเมทริกซ์ได้เช่นกัน

$$y = b_1x_1 + b_2x_2 + \cdots + b_px_p + e$$

$$y = \mathbf{b}'_{p \times 1} \mathbf{x}_{p \times 1} + e$$

- กรณีที่ตัวแปรตามมี q ตัว ทำนายจากตัวแปรอิสระ p ตัว สามารถขยายสมการได้ดังนี้

$$y_1 = b_{11}x_1 + b_{21}x_2 + \cdots + b_{p1}x_p + e_1$$

$$y_2 = b_{12}x_1 + b_{22}x_2 + \cdots + b_{p2}x_p + e_2$$

⋮

$$y_q = b_{1q}x_1 + b_{2q}x_2 + \cdots + b_{pq}x_p + e_q$$

$$\mathbf{y}_{q \times 1} = \mathbf{B}'_{p \times q} \mathbf{x}_{p \times 1} + \mathbf{e}_{q \times 1}$$

ค่าคาดหวังและความแปรปรวนร่วม



- ให้ x เป็นตัวแปรสุ่ม และ \mathbf{X} เป็นเวกเตอร์ตัวแปรสุ่ม
- ค่าคาดหวัง (Expectation) คือ หากสุ่มค่านี้มาเรื่อยๆ แล้ว ค่าเฉลี่ยในระยะยาวเท่ากับเท่าไร ซึ่งก็คือค่าเฉลี่ยในประชากร
 - $E(x_i) = E(x) = \mu$ กรณีที่มีตัวแปร x เพียงแค่ตัวเดียว
 - $E(\mathbf{x}_i) = E(\mathbf{X}) = \boldsymbol{\mu}$ กรณีที่มีตัวแปรหลายตัวแปร สุ่มออกมาแต่ละครั้ง จัดเก็บข้อมูลในรูปแบบของเวกเตอร์
- ให้ c เป็นค่าคงที่ และ \mathbf{c} เป็นเวกเตอร์ของค่าคงที่ ที่มีจำนวนสมาชิกเท่ากับสมาชิกของ \mathbf{x}
- ให้ $y = x + c$ และ $\mathbf{y} = \mathbf{X} + \mathbf{c}$ แล้ว
 - $E(y) = E(x + c) = E(x) + E(c) = \mu + c$
 - $E(\mathbf{y}) = E(\mathbf{X} + \mathbf{c}) = E(\mathbf{X}) + E(\mathbf{c}) = \boldsymbol{\mu} + \mathbf{c}$

ค่าคาดหวังและความแปรปรวนร่วม

- ให้ $z = c\mathbf{x}$ และ $w = \mathbf{c}'\mathbf{x} = c_1x_1 + c_2x_2 + \cdots + c_px_p$ แล้ว
 - $E(z) = E(cx) = cE(x) = c\mu$
 - $E(w) = E(\mathbf{c}'\mathbf{x}) = \mathbf{c}'\boldsymbol{\mu}$
- ในกรณีตัวแปรเดียว การหาความแปรปรวน (Variance) คือ หากสุ่มค่านี้มาเรื่อยๆ แล้ว ค่าความแปรปรวนในระยะยาวเท่ากับเท่าไร ซึ่งก็คือค่าความแปรปรวนในประชากร
 - $Var(x_i) = Var(x) = \sigma^2$
 - $Var(x + c) = Var(x) = \sigma^2$
 - $Var(cx) = c^2Var(x) = c^2\sigma^2$

ค่าคาดหวังและความแปรปรวนร่วม

- ในกรณีตัวแปรหลายตัวแปร การหาความแปรปรวน (Variance) จะถูกมองเป็นการหาความแปรปรวนร่วม (Covariance) คือ หากสุ่มค่านี้มาเรื่อยๆ แล้ว ค่าความแปรปรวนร่วมในระยะยาวเท่ากับเท่าไร ซึ่งก็คือค่าความแปรปรวนร่วมในประชากร
 - $Cov(\mathbf{x}_i) = Cov(\mathbf{x}) = \Sigma$
 - $Cov(\mathbf{x} + \mathbf{c}) = Cov(\mathbf{x}) = \Sigma$
 - $Cov(\mathbf{c}'\mathbf{x}) = \mathbf{c}'\Sigma\mathbf{c}$
- บางครั้ง อาจใช้ $Var()$ แทน $Cov()$ หากด้านในเป็นเวกเตอร์ ก็เข้าใจได้ว่าเป็นการหาความแปรปรวนร่วม

ค่าคาดหวังและความแปรปรวนร่วม

- ให้ x และ y เป็นตัวแปรสุ่ม (Random Variable) แล้ว
 - $E(x + y) = E(x) + E(y)$
 - $Var(x + y) = Var(x) + Var(y) + 2Cov(x, y)$
 - หาก x และ y ไม่สัมพันธ์กันแล้ว $Var(x + y) = Var(x) + Var(y)$
- ให้ \mathbf{x} และ \mathbf{y} เป็นเวกเตอร์ตัวแปรสุ่มแล้ว
 - $E(\mathbf{x} + \mathbf{y}) = E(\mathbf{x}) + E(\mathbf{y})$
 - $Cov(\mathbf{x} + \mathbf{y}) = Cov(\mathbf{x}) + Cov(\mathbf{y}) + Cov(\mathbf{x}, \mathbf{y}) + Cov(\mathbf{y}, \mathbf{x})$
 - ถ้า \mathbf{x} และ \mathbf{y} ไม่สัมพันธ์กัน กล่าวคือ $Cov(\mathbf{x}, \mathbf{y}) = Cov(\mathbf{y}, \mathbf{x}) = \mathbf{0}$ แล้ว
 $Cov(\mathbf{x} + \mathbf{y}) = Cov(\mathbf{x}) + Cov(\mathbf{y})$

ฟังก์ชันเพิ่มเติม

- Trace คือ ผลรวมของสมาชิกแนวทแยง

$$\text{tr}(\mathbf{A}_{1 \times 1}) = \text{tr}([1]) = 1$$

$$\text{tr}(\mathbf{B}_{2 \times 2}) = \text{tr}\left(\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}\right) = 2 + 3 = 5$$

$$\text{tr}(\mathbf{C}_{3 \times 3}) = \text{tr}\left(\begin{bmatrix} 5 & 2 & -7 \\ 2 & 12 & 5 \\ -7 & 5 & 11 \end{bmatrix}\right) = 5 + 12 + 11 = 28$$

ฟังก์ชันเพิ่มเติม

- Determinant เป็นค่าหนึ่งทางเมทริกซ์ที่ใช้ในการหาอินเวอร์ซ
- ในเชิงสถิติ จะหา Determinant ของเมทริกซ์ความแปรปรวนเป็นหลัก ค่า Determinant ของเมทริกซ์จะออกมาเป็นคะแนนเดี่ยว (Scalar) บอกความแปรปรวนแผ่ขยาย (Generalized Variance) ของตัวแปรทั้งหมด
- ถ้าตัวแปรสัมพันธ์กันมาก ความแปรปรวนแผ่ขยายจะต่ำ แต่หากตัวแปรไม่สัมพันธ์กันเลย ความแปรปรวนแผ่ขยายจะสูง

```
> S1 <- matrix(c(1, 0, 0,
+               0, 1, 0,
+               0, 0, 1), 3, 3, byrow=TRUE)
> det(S1)
[1] 1
> S2 <- matrix(c(1, 0.2, 0.3,
+               0.2, 1, -0.1,
+               0.3, -0.1, 1), 3, 3, byrow=TRUE)
> det(S2)
[1] 0.848
> S3 <- matrix(c(1, 0.5, 0.6,
+               0.5, 1, 0.7,
+               0.6, 0.7, 1), 3, 3, byrow=TRUE)
> det(S3)
[1] 0.32
> S4 <- matrix(c(1, 0.95, 0.95,
+               0.95, 1, 0.95,
+               0.95, 0.95, 1), 3, 3, byrow=TRUE)
> det(S4)
[1] 0.00725
```

ยิ่งตัวแปรไม่สัมพันธ์กัน ค่า **determinant**
ของเมทริกซ์สหสัมพันธ์ยิ่งสูง

หลักการความเป็นไปได้สูงสุด

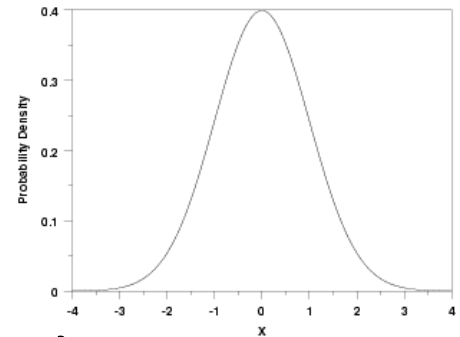
- สมมติว่ามีตัวแปรหนึ่ง เก็บข้อมูลจาก 4 คน มีค่า 2, 3, 4, และ 5 เชื่อว่าทั้ง 4 คนนี้ ถูกสุ่มมาจากประชากรที่มีการกระจายเป็นโค้งปกติ คะแนนแต่ละคนเป็นอิสระจากกันและมีการกระจายเป็นโค้งปกติเหมือนกัน (Independently and Identically Distributed, i.i.d.) การกระจายโค้งปกติใดที่เหมาะสมกับข้อมูลทั้ง 4 คนนี้

- สมการการกระจายโค้งปกติ เป็นดังนี้

$$f(x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}$$

- กล่าวคือ จะหาค่า μ และ σ ที่เหมาะสมกับข้อมูล 4 คน (2, 3, 4, 5) นี้

- $f(x)$ เป็นฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function; PDF)



หลักการความเป็นไปได้สูงสุด

- หลักการความเป็นไปได้สูงสุด คือ หาค่า μ และ σ ที่ทำให้ค่าด้านล่างมีค่าสูงสุด

จาก i.i.d

$$f(x_1, x_2, x_3, x_4) = f(x_1)f(x_2)f(x_3)f(x_4) = \frac{\exp\left(-\frac{1}{2}\left(\frac{2-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2}\left(\frac{3-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2}\left(\frac{4-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2}\left(\frac{5-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}$$

- ค่า $f(x)$ อาจเรียกว่าค่าความเป็นไปได้ (Likelihood) และหลักการที่ทำให้ค่าความเป็นไปได้สูงสุด จะเรียกว่าค่าความเป็นไปได้สูงสุด (Maximum Likelihood)
- ค่าความเป็นไปได้ จะมีคุณลักษณะคล้ายๆ ความน่าจะเป็น (Probability) ที่มีค่า 0 ถึง 1 และค่าความน่าจะเป็นจากหลายๆ เหตุการณ์ที่เป็นอิสระกัน จะหาได้โดยการคูณกัน เช่น
$$p(x \cap y) = p(x)p(y)$$
- ค่าความเป็นไปได้รวม จึงเป็นการคูณกันของความเป็นไปได้แต่ละตัว

หลักการความเป็นไปได้สูงสุด

- แต่ในคอมพิวเตอร์ จะไม่สามารถจัดการทศนิยมหลายๆ ตำแหน่งได้ง่าย และการคูณจะเพิ่มความยุ่งยากในการคำนวณ จึงนำ \log มาใส่ค่าความเป็นไปได้ เพื่อความสะดวกในการคำนวณ

$$\log f(x) = -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 - \log(\sigma\sqrt{2\pi})$$

- เมื่อใส่ \log แล้ว ผลคูณจะกลายเป็นผลรวม การหาค่าต่ำสุดของ \log -likelihood จะเท่ากับ การหาค่าสูงสุดของ likelihood

$$\log f(x_1, x_2, x_3, x_4) = -\frac{1}{2} \left(\frac{2 - \mu}{\sigma} \right)^2 - \log(\sigma\sqrt{2\pi}) - \frac{1}{2} \left(\frac{3 - \mu}{\sigma} \right)^2 - \log(\sigma\sqrt{2\pi}) - \frac{1}{2} \left(\frac{4 - \mu}{\sigma} \right)^2 - \log(\sigma\sqrt{2\pi}) - \frac{1}{2} \left(\frac{5 - \mu}{\sigma} \right)^2 - \log(\sigma\sqrt{2\pi})$$

หลักการความเป็นไปได้สูงสุด

- หากค่าผลรวม log-likelihood ต่ำสุดเป็นดังนี้

```
> ll <- function(x, vec) {  
+   mu <- x[1]  
+   sigma <- x[2]  
+   -sum(-0.5*((vec-mu)/sigma)^2 - log(sigma*sqrt(2*pi)))  
+ }  
> optim(c(3.5, 1), ll, vec=c(2,3,4,5))  
$par  
[1] 3.500263 1.117695  
  
$value  
[1] 6.122042
```

ค่าเฉลี่ยประมาณค่าได้ **3.5**

ค่าส่วนเบี่ยงเบนมาตรฐาน (แบบหารด้วย N) ได้ค่า **1.18**

หลักการความเป็นไปได้สูงสุด

- ค่า Likelihood ของการกระจายแบบโค้งปกติที่มีหลายตัวแปร (Multivariate Normal Distribution; MVN) เป็นดังนี้

$$f(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}}$$

- ส่วน log-likelihood ของ MVN เป็นดังนี้

$$\log f(\mathbf{x}) = -\frac{1}{2}(\ln(|\boldsymbol{\Sigma}|) + (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) + p \ln(2\pi))$$

- หาฟังก์ชันความหนาแน่นที่เหมาะสมกับข้อมูล โดยการหา $\boldsymbol{\mu}$ และ $\boldsymbol{\Sigma}$ ผ่าน Log-likelihood

หลักการความเป็นไปได้สูงสุด

- หากค่าของเวกเตอร์ค่าเฉลี่ย $\boldsymbol{\mu}$ และเมทริกซ์ความแปรปรวนร่วม $\boldsymbol{\Sigma}$ ที่เหมาะสมกับข้อมูล N คน กล่าวคือ $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ ได้ โดยให้ผลรวมของ log-likelihood ต่ำที่สุด

$$\log f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \sum_{i=1}^N \left[-\frac{1}{2} (\ln(|\boldsymbol{\Sigma}|) + (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) + p \ln(2\pi)) \right]$$

$$\log L = -\frac{N}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^N [\mathbf{y}_i - \boldsymbol{\mu}]' \boldsymbol{\Sigma}^{-1} [\mathbf{y}_i - \boldsymbol{\mu}] - \frac{Np}{2} \log(2\pi)$$

- หากตัวอย่างทั้งหมดไม่มีค่าสูญหาย สมการ log-likelihood จะเป็นดังนี้

$$\log L = -\frac{Np}{2} \log(2\pi) - \frac{N}{2} \log |\boldsymbol{\Sigma}| - \frac{N}{2} \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \frac{N}{2} [\mathbf{m}_y - \boldsymbol{\mu}]' \boldsymbol{\Sigma}^{-1} [\mathbf{m}_y - \boldsymbol{\mu}]$$

หลักการความเป็นไปได้สูงสุด

- โดยปกติ ค่า \mathbf{m}_y และ \mathbf{S} ที่คำนวณได้จากกลุ่มตัวอย่าง จะเป็นค่าที่มีความเป็นไปได้สูงสุด

$$\log L_S = -\frac{Np}{2} \log(2\pi) - \frac{N}{2} \log|\mathbf{S}| - \frac{1}{2} \sum_{i=1}^N [\mathbf{y}_i - \mathbf{m}_y]' \mathbf{S}^{-1} [\mathbf{y}_i - \mathbf{m}_y]$$

$$\log L_S = -\frac{Np}{2} \log(2\pi) - \frac{N}{2} \log|\mathbf{S}| - \frac{N}{2} p$$

- โมเดลที่ประมาณค่าเฉลี่ยและความแปรปรวนร่วมทั้งหมด จะเรียกว่าโมเดลแบบอิ่มตัว (Saturated Model) จะมีค่าความเป็นไปได้สูงที่สุด

หลักการความเป็นไปได้สูงสุด

- อย่างไรก็ตาม ใน SEM หรือสถิติอื่น จะสร้างค่าพารามิเตอร์ ที่จะบอกว่า ค่า m_y และ S ที่เป็นไปได้จะเป็นอย่างไร (เช่น กำหนดว่าตัวแปรทั้งหมดถูกอธิบายจากองค์ประกอบเดียว) ค่าเฉลี่ยและความแปรปรวนร่วมที่คำนวณได้จากโมเดลจะเรียกว่า ค่าเฉลี่ยและความแปรปรวนร่วมที่ได้จากโมเดล (Model-implied Means and Covariance Matrix)
- Maximum Likelihood (ML) คือ การค่าพารามิเตอร์ในโมเดล เพื่อทำให้ค่า log-likelihood มีค่าสูงที่สุด
 - ค่า log-likelihood ที่ได้จากโมเดล มักมีค่าต่ำกว่าที่ได้จากโมเดลอิมตัว

หลักการความเป็นไปได้สูงสุด

- ตัวอย่าง

- ข้อมูลของ Holzinger and Swineford (1939) เลือกเฉพาะตัวแปร x_1 และ x_2
- นำคะแนน x_1 มาบวกด้วยค่าคงที่ 1.1
- ค่าเฉลี่ยของ x_1 และ x_2 จะเท่ากับ 6.035 และ 6.088 ตามลำดับ
- นักวิจัยอาจต้องการเปรียบเทียบว่า ค่าเฉลี่ยของสองตัวแปรนี้แตกต่างกันหรือไม่ ด้วย dependent t -test พบว่า $M_1 - M_2 = -0.052, t(300) = -0.65, p = .51$

```
t_res <- t.test(dat3[,1], dat3[,2], paired = TRUE)
t_res
```

```
##
## Paired t-test
##
## data: dat3[, 1] and dat3[, 2]
## t = -0.65246, df = 300, p-value = 0.5146
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -0.2099241 0.1053837
## sample estimates:
## mean difference
## -0.05227021
```

หลักการความเป็นไปได้สูงสุด

- สามารถใช้สมการความเป็นไปได้ ในค่า \mathbf{m}_y และ \mathbf{S} ที่ทำให้เกิดความเป็นไปได้สูงสุด

$$\log L_S = -\frac{Np}{2} \log(2\pi) - \frac{N}{2} \log|\mathbf{S}| - \frac{1}{2} \sum_{i=1}^N [\mathbf{y}_i - \mathbf{m}_y]' \mathbf{S}^{-1} [\mathbf{y}_i - \mathbf{m}_y]$$

- พบว่า ค่า $m_1 = 6.04, m_2 = 6.09,$
 $s_1^2 = 1.36, s_2^2 = 1.38, s_{12} = 0.41$
- $\log L_S = -935.03$
- ในสมการนี้มีพารามิเตอร์ 5 ตัว

```
l12 <- function(x, data) {
  my <- matrix(c(x[1], x[2]), 2, 1)
  sy <- matrix(c(x[3], x[5], x[5], x[4]), 2, 2)
  invsy <- solve(sy)
  detsy <- det(sy)
  p <- 2
  n <- nrow(data)

  val3 <- 0
  for (i in 1:n) {
    diff <- matrix(data[i, ], p, 1) - my # vector length p
    val3 <- val3 + t(diff) %*% invsy %*% diff
  }
  val3 <- val3
  val2 <- n * log(detsy)
  val1 <- n * p * log(2 * pi)
  0.5 * (val1 + val2 + val3)
}
dat2 <- as.matrix(HolzingerSwineford1939[c("x1", "x2")])
dat3 <- dat2
dat3[,1] <- dat2[,1] + 1.1
optim(c(6.0, 6.1, 1.4, 1.4, 0.41), l12, data=dat3)
```

```
## $par
## [1] 6.0354824 6.0876296 1.3578378 1.3818295 0.4071527
##
## $value
## [1] 935.0332
##
## $counts
## function gradient
##      160      NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

หลักการความเป็นไปได้สูงสุด

- สมมติว่านักวิจัยลองเปรียบเทียบกับโมเดลที่ค่าเฉลี่ยของทั้งสองตัวแปรเท่ากัน กล่าวคือประมาณค่า 4 พารามิเตอร์ คือ m (ค่าเฉลี่ยรวมทั้งสองตัวแปร), s_1^2 , s_2^2 , s_{12} สามารถใช้หลักการความเป็นไปได้สูงสุดหาได้เช่นกัน โดยจำกัดการหาแค่ 4 พารามิเตอร์
- พบว่า ค่า $m = 6.06$, $s_1^2 = 1.36$,
 $s_2^2 = 1.38$, $s_{12} = 0.41$
- $\log L_S = -935.25$
- ในสมการนี้มีพารามิเตอร์ 4 ตัว

```
114 <- function(x, data) {  
  mu <- rep(x[1], 2)  
  Sigma <- matrix(c(x[2], x[4], x[4], x[3]), 2, 2)  
  
  # ensure positive-definite covariance  
  if (det(Sigma) <= 0) return(1e10)  
  
  invSigma <- solve(Sigma)  
  n <- nrow(data)  
  p <- ncol(data)  
  
  Z <- sweep(data, 2, mu) # center data  
  quad <- sum(diag(invSigma)%%crossprod(Z)) # ef  
  
  val1 <- n * p * log(2*pi)  
  val2 <- n * log(det(Sigma))  
  val3 <- quad  
  
  nll <- 0.5 * (val1 + val2 + val3) # negative Log-  
  return(nll)  
}  
optim(c(6.0, 1.4, 1.4, 0.41), 114, data=dat3)  
  
## $par  
## [1] 6.0615330 1.3591045 1.3826362 0.4066851  
##  
## $value  
## [1] 935.2466  
##  
## $counts  
## function gradient  
##      159      NA  
##  
## $convergence  
## [1] 0  
##  
## $message  
## NULL
```

หลักการความเป็นไปได้สูงสุด

- $H_1: \mu_1 \neq \mu_2$: ใช้ 5 พารามิเตอร์ ได้ $\log L_S = -935.03$
- $H_0: \mu_1 = \mu_2$: ใช้ 4 พารามิเตอร์ ได้ $\log L_S = -935.25$
- โมเดลแรกเรียกโมเดลเต็ม (Full Model) และโมเดลที่สองเรียกโมเดลที่ถูกจำกัด (Restricted Model)
- ในการตรวจสอบความเป็นไปได้สูงสุด ถ้าโมเดลทั้งสองซ้อนกัน (อธิบายในภายหลัง) สามารถนำมาเปรียบเทียบว่าโมเดลทั้งสองแตกต่างกันอย่างมีนัยสำคัญหรือไม่ ผ่านการทดสอบสัดส่วนความเป็นไปได้ (Likelihood Ratio Test) หรือการทดสอบความเบี่ยงเบน (Deviance Test): $D = -2 \times (LL_0 - LL_1)$
- จากตัวอย่างนี้ ค่าเท่ากับ $-2(-935.25 + 935.03) = 0.427$

หลักการความเป็นไปได้สูงสุด

- ค่า D นี้ หากกลุ่มตัวอย่างขนาดใหญ่ และโมเดลที่จำกัดถูกต้อง การกระจายของค่า D จะเป็นการกระจาย Chi-square ที่ df เท่ากับความแตกต่างของจำนวนพารามิเตอร์ของทั้งสองโมเดล
- ในตัวอย่างนี้ ความแตกต่างค่าพารามิเตอร์เท่ากับ 1
- ดังนั้น $\chi^2(1) = 0.427, p = .51$
- ค่า p เท่ากับ Dependent t -test ค่าเฉลี่ยของทั้งสองตัวแปรไม่แตกต่างกันอย่างมีนัยสำคัญ
- หาก $N \rightarrow \infty$ แล้ว $\chi^2 = t^2$